이학박사학위논문

# Highly specific genome editing in human cells and plants using CRISPR systems

**2017년 2월**

서울대학교 대학원

화학부 생화학 전공

김 정 은

Abstract

# Highly specific genome editing in human cells and plants using CRISPR systems

Jungeun Kim

Department of Chemistry

The Graduate School

Seoul National University

Targeted genome engineering with RNA-guided engineered nucleases (RGENs) repurposed from the clustered regularly interspaced short palindromic repeat (CRISPR) systems has recently received significant attention because of their simple design and preparation. Although, RGENs-mediated genome editing has been successfully

performed in cultured cells and model organisms, developing a highly efficient and specific CRISPR-based genome editing platform is still a major challenge in this area.

In this study, I develop and optimize the new CRISPR-based genome editing platforms and investigate their target specificities. First, I present delivery of RGEN RNPs in plant cells. RGEN RNPs enabled DNA-free genome editing in plant cells at high frequencies up to 44%. Whole plants were successfully regenerated from gene-edited cells and no apparent off-target mutations were detected in these mutants. This method may reduce the likelihood of DNA insertion and potential off-target effects and alleviate the concerns related to genetically modified plants. Second, I optimize the Cpf1-mediated genome editing system and investigate its target specificity. CRISPR-Cpf1 system induced targeted genome modification efficiently in human cells. Mismatched crRNAs test and genome-wide off-target analysis showed that this nuclease yielded DNA cleavages at the target site in a highly specific manner. These results indicate that CRISPR-Cpf1 is suitable for precision genome editing and will be a good option for RNA-guided genome editing.

**Keywords :** Genome engineering, Programmable nucleases, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), RNA-guided Endonucleases (RGENs), Cpf1

**Student Number : 2012-23038**

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| Cas | CRISPR-associated |
| CRISPR | Clustered regularly interspaced palindromic repeats |
| crRNA | CRISPR RNA |
| Cpf1 | CRISPR from Prevotella and Francisella 1 |
| DSB | Double-strand break |
| HEK | Human embryonic kidney |
| HR | Homologous recombination |
| NHEJ | Non-homologous end-joining |
| PAM | Protospacer adjacent motif |
| PCR | Polymerase chain reaction |
| PEG | Polyethylene glycol |
| RFLP | Restriction fragment length polymorphism |
| RGEN | RNA-guided engineered nuclease |
| RNP | Ribonucleoprotein |
| sgRNA | Single-guide RNA |
| T7E1 | T7 endonuclease 1 |
| ZFN | Zinc-finger nuclease |
| TALEN | Transcription activator-like effector nuclease |

# Chapter 1. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins

# Introduction

Understanding the function of genes and non-coding elements is one of the principal challenges in basic researches. To do this, gene-targeting (Capecchi, 1989) and siRNAs have been traditionally used for knocking-out and regulating the genetic elements. Gene-targeting uses homologous recombination between chromosomal DNA sequence and exogenously introduced DNA sequence, yielding a genetic modification of target site. This method is powerful in some systems such as mouse embryonic stem cells and yeasts, but the efficiency of gene-targeting is generally low (10-6 to 10-7) in eukaryotic cells, hampering the widespread use of this tool. In case of siRNAs, siRNAs can only target the genetic elements transcribed, indicating that function of genetic elements such as DNA binding motifs for transcription factors can't be interrogated with this method. Besides, siRNAs induce a gene knockdown non-specifically in some cases (Fedorov et al., 2006; Jackson et al., 2006).

Recently, programmable nucleases have emerged as a game changer in the field of genome engineering (Kim and Kim, 2014). Programmable nuclease is a kind of artificial enzyme, which can produce DNA double strand breaks (DSBs) at the desired site in the genome. DSBs generated by these nucleases boost cellular DNA repair mechanisms such as error-prone non-homologous end joining (NHEJ) and homology-directed repair (HDR) in the presence of a donor DNA, yielding a targeted gene modification at a high frequency. The three

most commonly used programmable nucleases are zinc-finger nucleases (ZFNs) (Bibikova et al., 2002; Kim et al., 1996; Lee et al., 2010; Urnov et al., 2005; Urnov et al., 2010), transcription activator-like effector nucleases (TALENs) (Boch et al., 2009; Christian et al., 2010; Kim et al., 2013; Miller et al., 2011; Moscou and Bogdanove, 2009) and RNA-guided engineered nucleases (RGENs) repurposed from the bacterial clustered regularly interspaced short palindromic repeat (CRISPR)–Cas (CRISPR-associated) system (Cho et al., 2013a; Cong et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Jinek et al., 2012; Jinek et al., 2013; Mali et al., 2013).

ZFN is a fusion protein composed of zinc-finger protein (ZFP) and nuclease domain of Fok I type IIS restriction endonuclease (Kim et al., 1996). Target specificity of nuclease is determined by ZFP and nuclease domain of Fok I produces DNA cleavages at the site recognized by ZFP. ZFP consists of zinc-finger modules, and each zinc-finger recognizes a specific 3-bp of DNA sequence. In case of TALENs, DNA cleavage activity of these proteins is triggered from the nuclease domain of Fok I like ZFNs, but transcription activator-like effectors (TALEs) determine the target site instead of ZFP. TALEs recognize the DNA sequence at a 1-bp resolution (Boch et al., 2009; Moscou and Bogdanove, 2009), which makes it easy to design a target site, expanding the design density of programmable nucleases. Meanwhile, RGENs recognize and cleave the target DNA in an entirely different way. RGENs consist of Cas9 protein and guide RNA. Guide RNA recognizes a target site by Watson-Crick base pairing and DNA

cleavages at the target site are triggered by Cas9 protein (Jinek et al., 2012). Target sites of RGENs can be easily replaced by preparing a new guide RNA only, which makes the breaking down of barriers of genome engineering resulting in the CRISPR Craze in biological sciences (Pennisi, 2013).

In the field of plant science, several groups independently reported that RGEN-mediated genome editing could induce targeted genome modifications in plant genomes at a high frequency (Li et al., 2013; Nekrasov et al., 2013; Shan et al., 2013). Compared with classical approaches to genetic modifications in plants, including a transferring the whole gene, random mutagenesis with chemicals and radiation and breeding, CRISPR-based technology enables a precise genome engineering because RGENs cut a predetermined site and induce small indels that are indistinguishable from naturally occurring genetic variations. RGEN-mediated genome editing also enables the rapid alteration of plants. Genome-editing with RGENs in plant systems has been established as a powerful tool for interrogating the gene function in academic researches.

Now, RGEN-mediated genome editing has drawn attention from not only basic research, but plant biotechnology and agriculture as a tool for improving crops. RGENs make it possible to generate crops that have new traits more straightforwardly. But, there is an unclear issue related to a regulatory approval of genetically modified crops obtained by CRISPR technology. The U.S. Department of Agriculture (USDA)'s rule has so far focused on whether genome-edited plants

contain foreign DNA and the USDA has recently decided that genome-edited plants obtained by transient transfection of plasmid DNA which encodes an expression cassette of programmable nucleases are exempted from genetically modified organism (GMO) legislation (Jones, 2015; Pennisi, 2016; Waltz, 2016). In case of EU and other regions, they have postponed its own decision. However, delivery of non-integrating plasmids into plant cells still has a matter of debate. In human cell experiments, several groups reported that a partial sequence of plasmid DNA could be inserted at the on-target and off-target sites of RGEN (Cradick et al., 2013; Fu et al., 2013). To rule out a little possibility of DNA insertion, delivery of preassembled CRISPR-Cas9 ribonucleproteins (Kim et al., 2014b) in plant cells might be a good approach to solve this problem.

Here, I design a platform for plant genome editing using preassembled CRISPR-Cas9 ribonucleoproteins (Kim and Kim, 2016; Woo et al., 2015) and show that RGEN RNP induce targeted genome modification at a high frequency and whole plants are successfully regenerated from RGEN RNP-transfected protoplasts. I propose that this method may circumvent regulatory issues related to genetically modified crops and facilitate the widespread use of the novel technology in the real world.

# Materials and Methods

## 1. Cas9 protein and guide RNA preparation

Cas9 protein tagged with HA epitope and nuclear localization signal was purchased from ToolGen, Inc. (South Korea). Purified Cas9 protein was stored against 20 mM HEPES (pH 7.5), 150 mM KCl, 1mM DTT and 10% glycerol at -20℃. Concentration of cas9 protein was adjusted to 10 mg/ml for ready to use. DNA templates for in vitro transcription of sgRNA were generated by oligo extension using phusion polymerase (New England BioLabs) and T7 RNA polymerase (New England BioLabs) was used to run-off in vitro transcription. For eliminating in vitro transcription template, DNase1 was treated to the reaction mixture after transcription. Transcribed sgRNAs were purified with MG PCR Product purification SV (Macrogen) and quantified by spectrometry.

## 2. Construction of RGEN-encoding plasmids

Cas9 expression plasmid for plant genome editing was modified from the human codon optimized p3s-CMV-hCas9 vector. For optimization, promoter was replaced to 35S promoter and BGH polyadenylation signal was replaced to 35S terminator sequence. sgRNA vector for protoplast test was designed to Arabidopsis U626 promoter-driven transcription. RGEN-encoding plasmids for protoplast

transfection were prepared with NucleoBond Xtra Midi kit (Macherey-Nagel).

## 3. Protoplast culture and isolation

Protoplasts were isolated as previously described from Arabidopsis, rice and lettuce. Initially, Arabidopsis (Arabidopsis thaliana) ecotype Columbia-0, rice (Oryza sativa L.) cv. Dongjin, and lettuce (Lactuca sativa L.) cv Cheongchima seeds were sterilized in a 70% ethanol, 0.4% hypochlorite solution for 15 min, washed three times in distilled water, and sown on 0.5× Murashige and Skoog solid medium supplemented with 2% sucrose. The seedlings were grown under a 16 h light (150 μmol m−2 s−1) and 8 h dark cycle at 25 °C in a growth room. For isolation of protoplasts, the leaves of 14 d Arabidopsis seedlings, the stem and sheath of 14 d rice seedlings, and the cotyledons of 7 d lettuce seedlings were digested with enzyme solution (1.0% cellulase R10, 0.5% macerozyme R10, 0.45 M mannitol, 20 mM MES [pH 5.7], CPW solution21) during incubation with shaking (40 r.p.m.) for 12 h at 25 °C in darkness and then diluted with an equal volume of W5 solution. The mixture was filtered before protoplasts were collected by centrifugation at 100g in a round-bottomed tube for 5 min. Re-suspended protoplasts were purified by floating on a CPW 21S (21% [w/v] sucrose in CPW solution, pH 5.8) followed by centrifugation at 80g for 7 min. The purified protoplasts were washed with W5 solution and pelleted by centrifugation at 70g for 5 min.

Finally, protoplasts were re-suspended in W5 solution and counted under the microscope using a hemocytometer. Protoplasts were diluted to a density of 1 × 106 protoplasts/ml of MMG solution (0.4 M mannitol, 15 mM MgCl2, 4 mM MES [pH 5.7]). In the case of tobacco protoplasts, 3-week-old Nicotiana attenuata leaves grown in B5 media were digested with enzymes (1% cellulose R10, 0.25% macerozyme R10, 0.5 M Mannitol, 8 mM CaCl2, 5 mM MES [pH 5.7], 0.1% BSA) for 5 h at 25 °C in darkness. Subsequently, protoplasts were washed with an equal volume of W5 solution twice. To obtain intact protoplasts, N. attenuata protoplasts in W5 solution were applied to an equal volume of 21% sucrose gradient followed by swing-out centrifugation at 50g for 5 min. The intact protoplasts were re-suspended in W5 solution and stabilized at least for 1 h at 4 °C before PEG-mediated transfection.

## 4. Protoplast transfection

PEG-mediated RNP transfections were performed as previously described. Briefly, to introduce DSBs using an RNP complex, 1 × 105 protoplast cells were transfected with Cas9 protein (10–60 μg) premixed with in vitro–transcribed sgRNA (20–120 μg). Prior to transfection, Cas9 protein in storage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, and 10% glycerol) was mixed with sgRNA in 1× NEB buffer 3 and incubated for 10 min at room temperature. A mixture of 1 × 105 protoplasts (or 5 × 105 protoplasts in the case of lettuce)

re-suspended in 200 μl MMG solution was gently mixed with 5–20 μl of RNP complex and 210 μl of freshly prepared PEG solution (40% [w/v] PEG 4000; Sigma No. 95904, 0.2 M mannitol and 0.1 M CaCl2), and then incubated at 25 °C for 10 min in darkness. After incubation, 950 μL W5 solution (2 mM MES [pH 5.7], 154 mM NaCl, 125 mM CaCl2 and 5 mM KCl) was added slowly. The resulting solution was mixed well by inverting the tube. Protoplasts were pelleted by centrifugation at 100g for 3 min and re-suspended gently in 1 ml WI solution (0.5 M mannitol, 20 mM KCl and 4 mM MES (pH 5.7)). Finally, the protoplasts were transferred into multi-well plates and cultured under dark conditions at 25 °C for 24–48 h. Cells were analyzed one day after transfection.

## 5. Genomic DNA isolation

Genomic DNA of protoplast and microcalli was isolated with DNeasy Plant Mini Kit (Qiagen). Genomic DNA of cotyledon derived from T1 plantlet was isolated with ZR Plant/Seed DNA MiniPrep (Zymo Research).

## 6. T7E1 assay

RGEN target site was amplified from purified genomic DNA by PCR. The target site contained-amplicon was denatured at 95℃ and cooled down to a room temperature slowly using a thermal cycler. Reannealed PCR amplicon was incubated with T7E1 (ToolGen, Inc.) in

1× NEB buffer 2 condition for 30 min at 37'C. The reaction mixture was analyzed by gel electrophoresis. DNA band intensities were calculated using densitometry.

## 7. Targeted deep sequencing

RGEN-target site and potential off-target sites were amplified from genomic DNA. Indices and sequencing adaptors were added by additional PCR. Paired-end sequencing was performed with a prepared amplicon library and MiSeq reagent kit v2 (300-cycles) was used for running. Raw fastq files were joined and indel reads were counted using a bioinformatic tool.

## 8. RGEN-RFLP

RGEN-RFLP assay was performed as previously described. In brief, 300 ng of PCR product was incubated with cas9 protein (1 μg) and sgRNA (750 ng) in a reaction volume of 10 μl at 37 °C for 60 min. NEB buffer 3 was used for RGEN-mediated cleavage. 4 μg of RNase A was used to remove the sgRNA after in vitro cleavage reaction. The reaction mixture was purified MG PCR Product purification SV (Macrogen) and purified sample was analyzed by agarose gel electrophoresis. DNA band intensities were measured by a densitometry.

## 9. Protoplast regeneration

RNP-transfected protoplasts were re-suspended in 0.5× B5 culture medium supplemented with 375 mg/l CaCl2•2H2O,18.35 mg/l NaFe-EDTA, 270 mg/l sodium succinate, 103 g/l sucrose, 0.2 mg/l 2,4-dichlorophenoxyacetic acid (2,4-D), 0.3 mg/l 6-benzylaminopurine (BAP) and0.1 g/l MES. The protoplasts were mixed with a 1:1 solution of 0.5× B5 medium and 2.4% agarose to a culture density of 2.5 × 105 protoplasts/ml. The protoplasts embedded in agarose were plated onto 6-well plates, overlaid with 2 ml of liquid 0.5× B5 culture medium, and cultured at 25 °C in darkness. After 7 days, the liquid medium was replaced with fresh culture medium. The cultures were transferred to the light (16 h light [30 μmol m−2 s−1] and 8 h darkness) and cultured at 25 °C. After 3 weeks of culture, micro-calli grown to a few millimeters in diameter were transferred to MS regeneration medium supplemented with 30 g/l sucrose, 0.6% plant agar, 0.1 mg/l α-naphthalaneacetic acid (NAA), 0.5 mg/l BAP. Induction of multiple lettuce shoots was observed after about 4 weeks on regeneration medium.

**Table 1. Oligonucleotides used for in vitro transcription template.**

| | Sequence (5' to 3') |
|---|---|
| AOC-sgF | GAAATTAATACGACTCACTATAGCAAAAGACTGTCAATTCCCTGTTTTAGAGCTAGAAATAGCAAG |
| PHYB-sgF | GAAATTAATACGACTCACTATAGGCACTAGGAGCAACACCCAACGTTTTAGAGCTAGAAATAGCAAG |
| P450-sgF | GAAATTAATACGACTCACTATAGGCATATAGTTGGGTCATGGCAGTTTTAGAGCTAGAAATAGCAAG |
| DWD1-TS1-sgF | GAAATTAATACGACTCACTATAGGTGCATCGTCCAAGCGCACAGGTTTTAGAGCTAGAAATAGCAAG |
| DWD1-TS2-sgF | GAAATTAATACGACTCACTATAGGCTACGACGTCAGGTTCTACCGTTTTAGAGCTAGAAATAGCAAG |
| BRI1-TS1-sgF | GAAATTAATACGACTCACTATAGGTTTGAAAGATGGAAGCGCGGGTTTTAGAGCTAGAAATAGCAAG |
| BRI1-TS2-sgF | GAAATTAATACGACTCACTATAGGTGAAACTAAACTGGTCCACAGTTTTAGAGCTAGAAATAGCAAG |
| BIN2-sgF | GAAATTAATACGACTCACTATAGATCACAGTGATGCTCGTCAAGTTTTAGAGCTAGAAATAGCAAG |
| Universal sgR | AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC |

**Table 2. Oligonucleotides used for T7E1 assay.**

| | 1st PCR | | 2nd PCR | |
|---|---|---|---|---|
| Target | Forward (5' to 3') | Reverse (5' to 3') | Forward (5' to 3') | Reverse (5' to 3') |
| AOC | CGAGCTCAATGAACGTGACC | GATCAGAATGCAGAGTCCAGC | | ATGCAGAGTCCAGCCGTTAT |
| PHYB | TGGTTGTTTGCCATCACACT | GAAAAGCCTGAAAGGACGAA | | GCCTCCCCATTTGATTTCTT |
| P450 | GGAGCTGAACCACTTCATCC | CCCAGCACCTGCTTCACTAT | ACCCCAGGCCAATTCATG | GGGACAAAGATTCATGCAGCA |
| DWD1 | CCTTTTCTTTGTGGGGTGTG | TCCTTCTCCCTCTCCTCCTG | ATCTCGTGCCATCTCCATCC | |
| BRI1 | ATTTGGGCTGATCCTTGTTG | TGTTGAACACCTGAAACTTTGG | ACCAATTGGAAGCTGACTGG | CCATGCCAAAATCTGAAACC |

# Results

## 1. Targeted gene modification in plant protoplasts using RGEN-RNPs

### a. Targeted gene disruption in protoplasts of plants

In the previous studies, RGEN-mediated genome editing was achieved using delivery vectors such as A. tumefaciens and plasmid DNA (Li et al., 2013; Nekrasov et al., 2013; Shan et al., 2013). These methods have some possibility of insertion of DNA fragments at both on-target and off-target sites. To rule out the possibility of inserting recombinant DNA in the host genome, RGEN RNP-mediated plant genome editing system was designed (Figure 1). To test whether PEG-mediated transfection of RGEN RNP can induce targeted gene modification, sgRNA targeting the PHYTOCHROME B (PHYB) gene was designed and PEG-mediated transfection was performed (Figure 2). I measured the gene modification efficiency by the T7E1 assay (Kim et al., 2009) and targeted deep sequencing. I found that RGEN RNP induced targeted gene modification of the PHYB gene in Arabidopsis and indel rates were increased in a dose dependent manner. Mutation rates were saturated when 30 μg of cas9 protein and 60 μg of sgRNA were used.

Next, I investigated whether RGEN RNP can induce targeted

gene modification in protoplasts of other plants. The RGEN RNPs were incubated with protoplasts derived from tobacco (Nicotiana attenuata) and rice (Oryza sativa) in the presence of polyethylene glycol (PEG). I used both the T7E1 assay and targeted deep sequencing to measure mutation frequencies in transfected cells (Figure 3a). Indels were detected at the expected position upstream of an NGG protospacer-adjacent motif (PAM) (Figure 3b), with frequencies that ranged from 8.4% to 44%. These results show that RGEN RNP efficiently introduced targeted mutation of plant genome regardless of the source of protoplasts.

Two sgRNAs were designed at the BRASSINOSTEROID INSENSITIVE 1 (BRI1) gene and co-transfected into protoplasts of Arabidopsis to investigate whether the repair of two concurrent DSBs would result in targeted deletion of the intervening sequence. Genomic DNA of transfected cells was isolated and target site-contained region was amplified by PCR. The PCR product was cloned into TOPO cloning vector and sanger sequencing was performed. Sanger sequencing showed that a 223-bp DNA sequence was deleted in protoplasts (Figure 4). Notably, RGEN-induced mutations were detected 24 h after transfection, suggesting that RGENs cut target sites immediately after transfection and induce mutation before a full cycle of cell division is completed (Figure 4a).

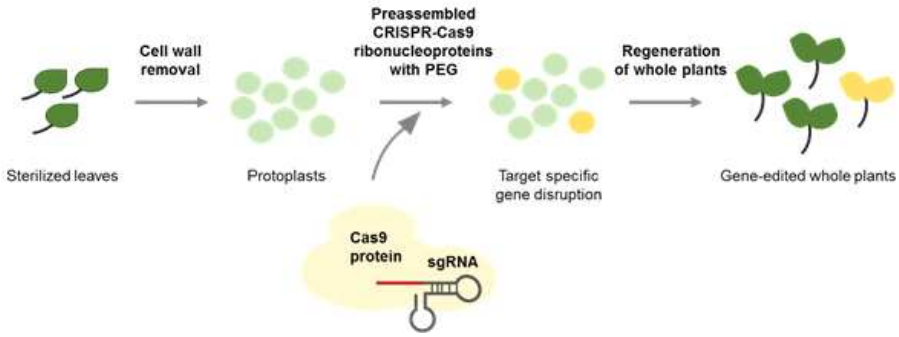



**Figure 1. Scheme for RGEN RNP-mediated genome editing in plant protoplasts.** Plant protoplasts prepared by enzymatic methods are transfected with preassembled CRISPR-Cas9 ribonucleoproteins using polyethylene glycol (PEG). Target specific gene disruption is introduced and genome-edited plants are regenerated from single mutant cells.

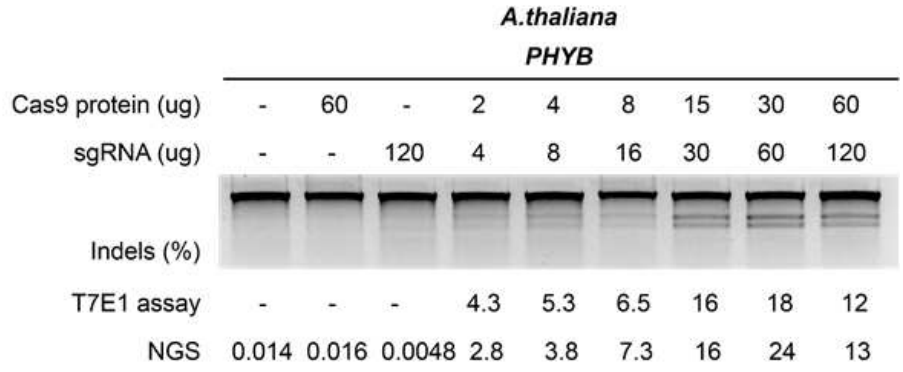
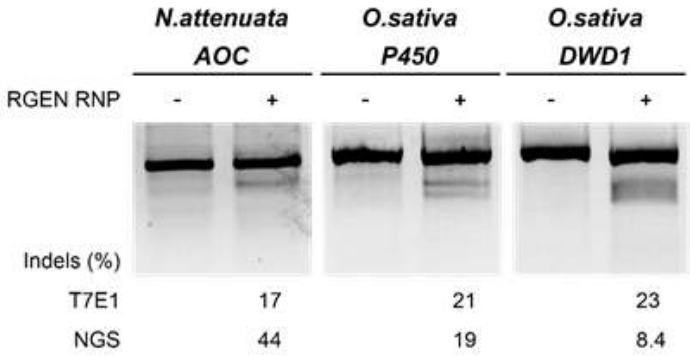



**Figure 2. RGEN RNP-mediated gene modification in plant protoplasts of Arabidopsis thaliana.** Purified Cas9 protein and PHYB gene-specific sgRNA were mixed and RGEN RNP complex were incubated with protoplasts derived from Arabidopsis thaliana. Indel rate were analyzed by T7E1 assay and targeted deep sequencing. Indel rates are increasing in a dose-dependent manner.

**a**




**b**

***AOC***

| | | | | |
|---|---|---|---|---|
| CAAAAGACTGTCAATTC-CCT**TGG** | WT | CATATAGTTGGGTCATG-GCA**TGG** | WT |
| CAAAAGACTGTCAATTCaCCTTGG | +1 | CATATAGTTGGGTCAT--GCATGG | -1 |
| CAAAAGACTGTCAATTCtCCTTGG | +1 | CATATAGTTGGGTC----GCATGG | -3 |
| CAAAAGACTGTCAATTCcCCTTGG | +1 | CATATAGTTGGGc-----GCATGG | -4 |
| CAAAAGACTGTCAATT--CCTTGG | -1 | CATATAGTTGGGT------CATGG | -5 |

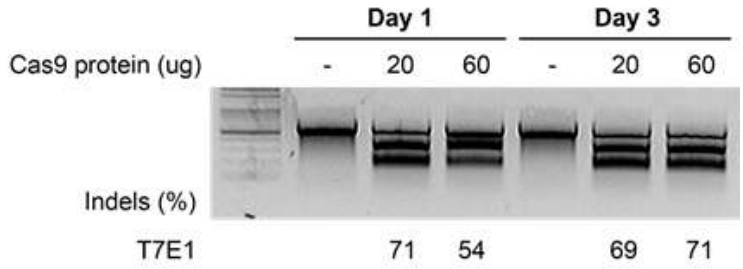
***DWD1*-TS1+TS2**

| | |
|---|---|
| TGCATCGTCCAAGCGCACAG**TGG**CCCGGCCTACGACGTCAGGTTCT-----ACC**CGG** | WT |
| TGCATCGTCCAAGCGC----------T-----ACCCGG | -29 |
| TGCATCGTCCAAGCGCACAGTGGCCCGGCCTACGACGTCAGGTTCT(ins)ACCCGG | +33 |

(With Hyeran Kim in Institute for Basic Science)

**Figure 3. RGEN RNP-mediated gene modification in plant protoplasts of Nicotiana attenuata and Oryza sativa.** (a) Mutation frequencies measured by the T7E1 assay and targeted deep sequencing. (b) Mutant DNA sequences induced by RGEN RNPs in plant cells. The RGEN target sequences are underlined and the PAM sequences are shown in bold. Insertion sequences are indicated by lowercase letter. WT, wild-type.

**a**




**b**

***BRI1***

TTTGAAAGATGGAAGCG-CGG**TGG.**(201bp).TGAAACTAAACTGGTCC-ACA**CGG** WT
TTTGAAAGATGGAAGCG----------C-ACACGG -223

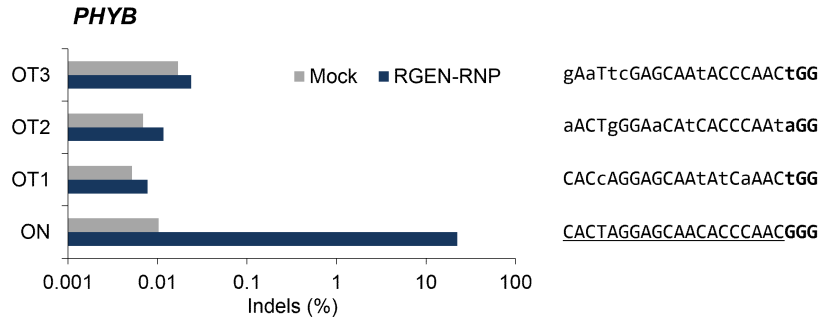
(By Seung Woo Cho in Seoul National University)

**Figure 4. A Time-course analysis of gene modification efficiency.** (a) A time-course analysis of genome editing of the BRI1 gene in A. thaliana protoplasts. Indel rate was analyzed by the T7E1 assay. (b) Mutant sequence of dual sgRNAs-mediated large deletion. The RGEN target sequences are underlined and the PAM sequences are shown in bold.

## b. Analysis of off-target effects

Next, I investigated whether RGEN RNPs induce DNA cleavages at sites highly homologous to the on target site. Potential off-target sites of the PHYB gene –specific sgRNAs were searched in the Arabidopsis genome using the Cas-OFFinder program (Bae et al., 2014) and off-target effects were validated by targeted deep sequencing. RGEN RNPs did not induce any indels above basal sequencing error levels at any sites (Figure 5) despite the high frequency of on-target indel rate. This result indicates that RGEN RNPs induce DNA cleavages of plant genome in a highly specific manner, in line with previous findings in human cells (Cho et al., 2014).

**a**




**b**

| Target gene | Sites | Sequence | Indel frequency (%) | | p-value |
|---|---|---|---|---|---|
| | | | Mock | RGEN RNP | |
| PhyB | ON | CACTAGGAGCAACACCCAACGGG | 0.010 | 22 | < 0.05 |
| | OT1 | CACcAGGAGCAAtAtCaAACTGG | 0.0052 | 0.0077 | 0.75 |
| | OT2 | aACTgGGAaCAtCACCCAAtAGG | 0.0069 | 0.012 | 0.33 |
| | OT3 | gAaTtcGAGCAAtACCCAACTGG | 0.017 | 0.024 | 0.37 |

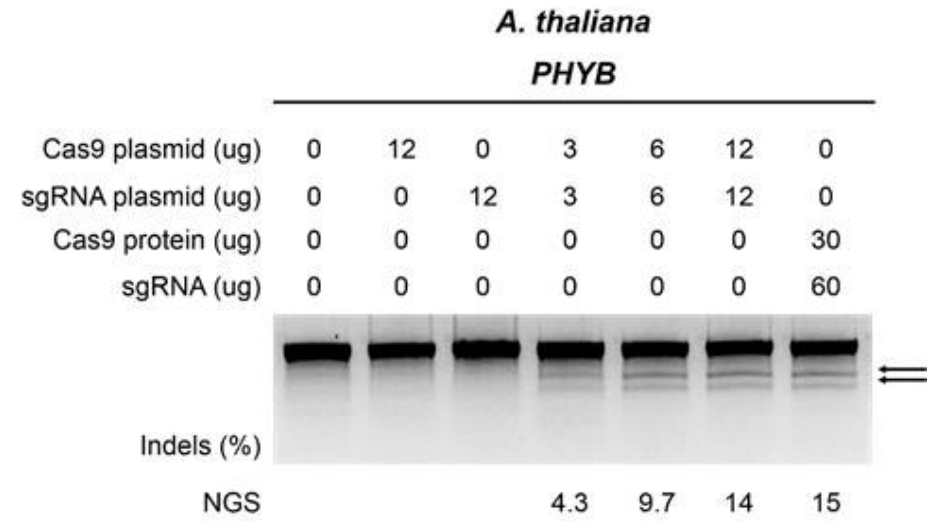
**Figure 5. Targeted deep sequencing of potential off-target sites of the PHYB and BRI1 gene-specific sgRNAs.** (a) Potential off-target sites of PHYB gene specific sgRNA were analyzed by targeted deep sequencing. The RGEN target sequences are underlined and the PAM sequences are shown in bold. Mismatched nucleotides are shown in lower case. (b) Summary of indel rates of PHYB gene on-target site and potential off-target sites.

## c. Comparison of the delivery methods of RGEN

In the previous study, RGEN-mediated plant genome editing was achieved by a transient transfection of plamid DNA-encoding Cas9 gene and sgRNAs (Li et al., 2013; Shan et al., 2013). To compare the genome-editing efficiency of RGEN RNP delivery to that of plasmid transfection, I designed RGEN-encoding plasmids optimized for plant systems. Cas9 expression plasmid was designed to transcribe cas9 gene under a 35S promoter and sgRNA vector was designed to Arabidopsis U626 promoter-driven transcription. Transfection was performed with RGEN-encoding plasmids and RGEN RNPs about the PHYB gene target site. Genomic DNA was isolated 72h post-transfection and T7E1 assay and targeted deep sequencing were performed to measure the gene modification efficiencies. I found that RGEN RNP-method induced mutations at a high frequency, comparable to those obtained with RGEN-encoding plasmids (Figure 6a). Notably, plasmid-derived sequences were found at the target site from raw NGS data obtained by transfection of RGEN-encoding plasmids, regardless of the amount of the plasmids used (Figure 6b). The frequencies of these insertions were low (0.06-0.14 % of total mutations) but are likely underestimated, because insertions of >50-bp sequences were excluded from deep-sequencing analysis. As expected, no plasmid DNA insertions were detected at the target site, when preassembled CRISPR-Cas9 ribonucleoproteins were used. These results indicate that RGEN RNP delivery is an effective method for genome engineering in plant systems

and has advantages for the issue about DNA integration into the host genome.

**a**




**b**

```
CACTAGGAGCAACACCC----------AACGGG    WT
CACTAGGAGCAACACCgTAGTTATCTGAACGGG   +10 (AmpR)

CACTAGGAGCAACACCC-----------AACGGG    WT
CACTAGGAGCAACACCCGGAGGACCGAAAACGGG   +11 (AmpR)

CACTAGGAGCAACACCC------------AACGGG    WT
CACTAGGAGCAACACCCTAGGAGCAACACccaacG   +12 (gRNA guide sequence)

CACTAGGAGCAACACCC---------------AACGGG    WT
CACTAGGAGCAACACCCCGCTCGCCGCAGCCGAACGaG   +15 (Cas9 plasmid backbone)

CACTAGGAGCAACACCC---------------------------------AACGGG    WT
CACTAGGAGCAACACCCAAGGCCTTTTGCTGGCCTTTTGCTCACATGTTCAACGGG   +33 (pUC origin)

CACTAGGAGCAACACCC-----------------------------------AACGGG    WT
CACTAGGAGCAACACCtGATGATCAGGTCCTTCTTCACCTCCTTGTAGCCCTAACGGG   +35 (cas9 sequence)

CACTAGGAGCAACACCC-----------------------------------AACGGG    WT
CACTAGGAGCAACACCCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACAACGGG   +36 (AmpR)

CACTAGGAGCAACACCC-------------------------------------------AACGGG    WT
CACTAGGAGCAACACCCGCCCTGATCAAGAAGTACCCCAAGCTGGAGAGCGAGTTCGTGTAACGGG   +43 (cas9 sequence)

CACTAGGAGCAACACCC-----------------------------------------------AACGGG    WT
CACTAGGAGCAACACCaGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCAACGGG   +45 (pUC origin)

CACTAGGAGCAACACCC-------------------------------------------------AACGGG    WT
CACTAGGAGCAACAACgATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGCAACGGG   +46 (AmpR)
```

**Figure 6. Mutation frequencies obtained with plasmid transfection and RNP delivery.** (a) Indel rates were analyzed by the T7E1 assay and targeted deep sequencing. Arrows indicate cleaved DNA bands. (b) Inserted sequences were obtained from raw NGS data. Inserted sequences were underlined and aligned with Cas9 and sgRNA-encoding plasmids. Origins of inserted sequences were indicated in parentheses. The PAM sequences are shown in bold and mismatched nucleotides are shown in lower case.

## 2. Targeted gene knockout and whole plant regeneration in lettuce using RGEN-RNPs

### a. Generation of BIN2 gene knockout lettuce

To test whether RGEN RNP methods could be used for generating a mutant plant, RGEN RNPs were transfected into protoplasts derived form lettuce (Lactuca sativa) to disrupt the lettuce homolog of the A. thaliana BRASSINOSTEROID INSENSITIVE 2 (BIN2) gene (Figure 7). BIN2 gene is a negative regulator in a brassinosteroid (BR) signaling pathway (Choe et al., 2002) and knocking-out of BIN2 gene is expected to lead lettuce to having more tolerance to various abiotic stresses (Koh et al., 2007). RGEN RNP-transfected protoplast were seeded onto 6-well plates, and maintained in regeneration medium. After 3 to 4 weeks of culture, single cell-derived microcalli were regenerated (Figure 8).

To analyze the genotype of microcalli, I performed RGEN RFLP analysis (Kim et al., 2014a). This analysis uses RGEN RNP complexes to cut the PCR amplicon harboring the target site. Cutting the RGEN target site in vitro leads to 100 % cleavage for wild-type clones, 50 % cleavage for monoallelic mutant clones, and no cleavage for heterozygous and homozygous biallelic mutant clones. 19 of 35 samples were identified as wild-type clones and the rest of the samples contained mono allelic mutation or bi-allelic mutation at the target site (Figure 9a). The overall mutation frequency in lettuce calli was 46%.

To identify the mutated sequence, clonal samples were subjected to targeted deep sequencing (Figure 9b,c). The number of base pairs deleted or inserted at the target site ranged from −9 to +1, consistent with mutagenic patterns observed in human cells (Cho et al., 2013a). No apparent mosaicism was detected in these clones, suggesting that the RGEN RNP cleaved the target site immediately after transfection and induced indels before cell division was completed.

Subsequently, whole plants were regenerated from genome-edited calli (Figure 10). Shoot formation was induced from protoplast-calli and root-induction was performed onto elongated shoots. Regenerated plantlets were transferred and grown in soil. These results show that whole plants were successfully regenerated from genome-edited protoplasts with RGEN RNP methods.

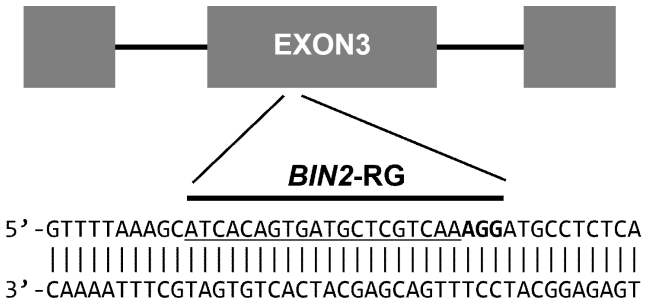
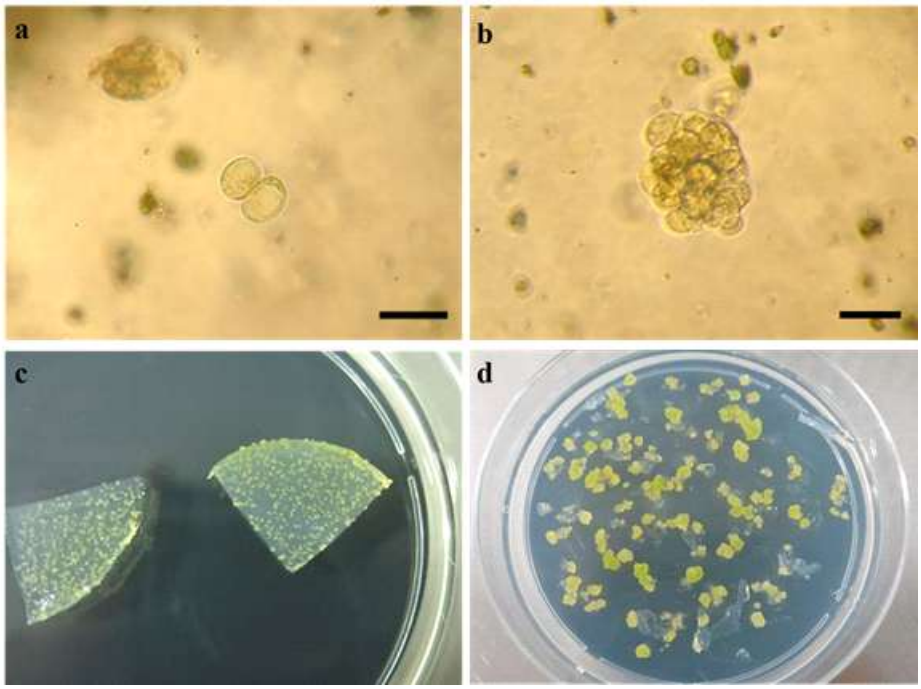


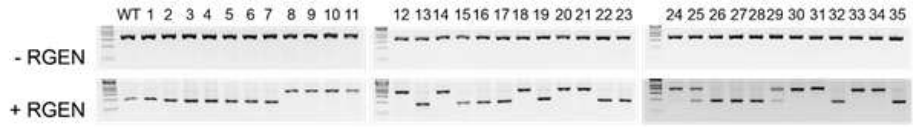

**Figure 7. The target sequence in the BIN2 gene.** The PAM sequence is shown in bold and the guide sequence is underlined.




(By Woo Je Wook in Advanced Institutes of Convergence Technology)

**Figure 8. Microcalli regenerated from the RGEN RNP-transfected cells.** (a) Cell division after 5 days of protoplast culture (Bar = 100 μm). (b) A multicellular colony of protoplast (Bar= 100 μm). (c) Agarose-embedded colonies after 4 weeks of protoplast culture. (d) Callus formation from protoplast-derived colonies.

**a**



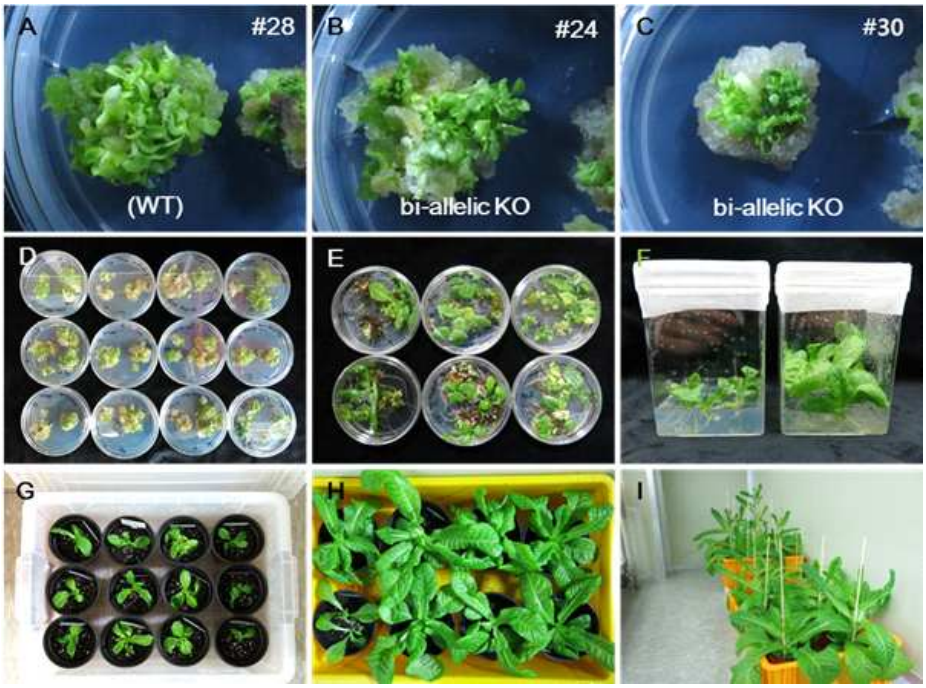

**b**

```
        ATCACAGTGATGCTCGT-CAAAGG WT
#08     ATCACAGTGATGCTCGTtCAAAGG +1
#09     ATCACAGTGATGCTCGTtCAAAGG +1
#10-A1  ATCACAGTGATGCT----CAAAGG -3
   -A2  ATCACAGTGATGCTCGTtCAAAGG +1
#11-A1  ATCACAGT----------CAAAGG -9
   -A2  ATCACAGTGATGCTCGTtCAAAGG +1

        ATCACAGTGATGCTCGT-CAAAGG WT
#12     ATCACAGTGATGCTCGTtCAAAGG +1
#14-A1  ATCACAGT----------CAAAGG -9
   -A2  ATCACAGTGATGCTCGTtCAAAGG +1
#18     ATCACAGTGATGCTCGTtCAAAGG +1
#20-A1  ATCACAGTGATGCTCGT---AAGG -2
   -A2  ATCACAGTGATGCTCGTtCAAAGG +1
#21     ATCACAGTGATGCTCGTtCAAAGG +10

        ATCACAGTGATGCTCGT-CAAAGG WT
#24-A1  ATCACAGTGATGCTCGTtCAAAGG +1
   -A2  ATCACAGTGATGCTCG--CAAAGG -1
#25-A1  ATCACAGTGATGCTCGTtCAAAGG +1
   -A2  ATCACAGTGATGCTCGT-CAAAGG WT
#29-A1  ATCACAGTG-------T-CAAAGG -7
   -A2  ATCACAGTGATGCTCGT-CAAAGG WT
#30     ATCACAGTGATGCTCGTtCAAAGG +1
#31     ATCACAGTGATGCTCGTtCAAAGG +1
#33     ATCACAGTGATGCTCGTtCAAAGG +1
#34     ATCACAGTGATGCTCGTtCAAAGG +1
```

**c**

| Genotype | Count | Frequency (%) |
|---|---|---|
| Wild type | 19 | 54 |
| Mono allele mutation | 2 | 5.7 |
| Bi-allele mutation (hetero) | 5 | 14 |
| Bi-allele mutation (homo) | 9 | 26 |
| Total number of mutant T0 plantlets | 16 | 46 |

**Figure 9. Genotyping of microcalli.** (a) Genotyping of microcalli with RGEN RFLP assay. No cleavage indicates biallelic mutant clone and 50 % cleavage indicates monoallelic mutant. WT clones show 100 % cleavage. (b) Genetic analysis of microcalli by targeted deep sequencing. The RGEN target sequence is underlined and the PAM sequence is shown in bold. Inserted sequences are shown in lower case. A1 indicates allele 1 and A2 indicates allele 2. (c) Summary of analyzed data of microcalli.




(By Woo Je Wook in Advanced Institutes of Convergence Technology)

**Figure 10. Whole plants regenerated from genome-edited calli.** (a-c) Organogenesis and shoot formation from protoplast-derived calli. Wild typd, #28. Heterozygous bi-allelic mutant, #24. Homozygous bi-allelic mutant, #30. (d) In vitro shoot proliferation and development. (e) Elongation and growth of shoots in MS culture medium free of PGR. (f) Root induction onto elongated shoots. (g) Acclimatization of plantlets. (h, i) Regenerated plants.

## b. Analysis of off-target effects in regenerated plantlets

To evaluate whether the BIN2-specific RGEN induced off target mutations in the lettuce genome, potential off-target sites of BIN2 gene-specific sgRNAs were searched in silico using the Cas-OFFinder program (Bae et al., 2014). Homologous sequences that differed from the on-target sequence by up to 5 nucleotides of mismatches were searched. The Legassy_V2 database was used as a reference genome sequence (Table 3). Genomic DNA was isolated from three BIN2-mutated plantlets and targeted deep sequencing libraries were generated from these samples by PCR. Some sites were not amplified successfully because of a poor quality of reference genome data. Indel rates of 92 sites were analyzed and no off-target mutations were detected at 91 homologous (Table 4), consistent with findings in human cells that off-target mutations induced by CRISPR RGENs are rarely found in a single cell-derived clone (Kim et al., 2015).

| | No. of mismatches to on-target site | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | Total |
| No. of potential off-target sites | 1 (on-target) | 0 | 1 | 4 | 27 | 349 | 382 |
| No. of sites with appropriate PCR primers | 1 | 0 | 1 | 3 | 24 | 72 | 101 |
| No. of sites amplified successfully | 1 | 0 | 1 | 3 | 22 | 65 | 92 |

**Table 3. Number of potential off-target sites of BIN2 gene-specific sgRNAs in the lettuce genome.** Potential RGEN off-target sites were identified in the lettuce genome using Cas-OFFinder (www.regenome.net). We used the Legassy_V2 database (Genebank: AFSA00000000.1) as the reference genome and identified homologous sequences that differed from on-target sequences by up to 5 nt. We chose a total of 92 sites and performed targeted deep sequencing. Some sites were excluded in this analysis because PCR primers couldn't be designed owing to a poor quality of reference genome data or because no amplicons were obtained using PCR.

| | | WT | T0-20 | T0-24 | T0-25 |
|---|---|---|---|---|---|
| Site name | Sequence | Indels(%) | Indels(%) | Indels(%) | Indels(%) |
| On-target | ATCACAGTGATGCTCGTCAAAGG | 0.021 | 99.912 | 99.867 | 45.042 |
| OT1 | ATCACAGTGcgGCTCGTCAAgGG | 0.022 | 0.039 | 0 | 0 |
| OT2 | caCACAGTGATGtTCGTCAAgGG | 0 | 0.014 | 0.024 | 0.013 |
| OT3 | ATacCAGgGATGCTCGTCAAtGG | 0 | 0 | 0 | 0 |
| OT4 | ATCAtAGTGATGCTCaTgAAgGG | 0.013 | 0.03 | 0.019 | 0 |
| OT5 | ATCACAtTGATGCTCtaCAtAGG | 0.023 | 0.033 | 0.029 | 0.012 |
| OT6 | ATaACAGaGAcGaTCGTCAAAGG | 0.029 | 0.03 | 0.014 | 0.027 |
| OT7 | ATCACAcTGATGCcCtaCAAAGG | 0.093 | 0.06 | 0.07 | 0.109 |
| OT8 | ATCACAtTGAgGCcCGaCAAAGG | - | - | - | - |
| OT9 | ATCACAcTGATGCaCtaCAAAGG | 0.057 | 0.037 | 0.023 | 0.077 |
| OT10 | caCACAGTGATGtTCaTCAAAGG | 0.635 | 0.715 | 0.663 | 0.145 |
| OT11 | ATgACAaTtATGCTCtTCAAAGG | 0.250 | 0 | 0.102 | 0 |
| OT12 | ATCAaAGTGcTcCTCGTgAAAGG | 0 | 0 | 0 | 0 |
| OT13 | taCACAaTGtTGCTCGTCAAcGG | 0.013 | 0 | 0 | 0.012 |
| OT14 | gcCACAGTGATGaTCGTCgAcGG | 0 | 0 | 0 | 0.013 |
| OT15 | ATatCAGgGATGCTCGcCAAtGG | 0 | 0 | 0 | 0 |
| OT16 | AaatCAGTGATcCTCGTCAAcGG | 0 | 0 | 0 | 0.012 |
| OT17 | ATggCAGTGATGgTCGTgAAgGG | 0 | 0.045 | 0.08 | 0.1 |
| OT18 | cTCAgAGTGtTGCTCtTCAAtGG | 0 | 0.01 | 0 | 0 |
| OT19 | ATCACAGaGATGCTCcaaAAtGG | 0.074 | 0.033 | 0.068 | 0.068 |
| OT20 | ATCAagGTtATtCTCGTCAAgGG | 0 | 0.009 | 0 | 0 |
| OT21 | AgCACAGTGAgGCTtGTCgAgGG | 0 | 0 | 0 | 0 |
| OT22 | ATatCAagGATGCTCGTCAAtGG | 0 | 0 | 0 | 0 |
| OT23 | tTCcCAGaGATGCTCtTCAAgGG | 0.024 | 0.05 | 0 | 0.035 |
| OT24 | gTCACAtTGATGCTCaTCAtgGG | 0 | 0 | 0.157 | 0 |
| OT25 | ATCACAGaGATGtTCaTCAtcGG | 0.022 | 0 | 0.013 | 0 |
| OT26 | ATCAaAaTGAgGCTCGaCAAcGG | - | - | - | - |
| OT27 | ATaACAaTGAaGCTCGTtAAtGG | 0 | 0 | 0 | 0 |
| OT28 | ATatCAGgGATGCTCaTCAAtGG | 0 | 0.011 | 0 | 0.017 |
| OT29 | ATCAtAtTGAaGCaCtTCAAAGG | 0.029 | 0.019 | 0 | 0.036 |
| OT30 | cTCACAtTGATGCaCtaCAAAGG | 0.069 | 0.055 | 0.073 | 0.097 |
| OT31 | tcCACAaTGATGCaCtTCAAAGG | 0.023 | 0 | 0 | 0.012 |
| OT32 | cTCACAaTGtTGCTCtaCAAAGG | - | - | - | - |
| OT33 | ATgACAaTGAaGCTCGTaAtAGG | 0 | 0 | 0.018 | 0 |
| OT34 | cTCtCAGTGgTGCTgGTCgAAGG | 0 | 0 | 0 | 0.029 |
| OT35 | ATCACAcTtATaCTCGaCAgAGG | 0 | 0 | 0.054 | 0.018 |
| OT36 | cTCACAGTGAgGCTttTaAAAGG | 0.16 | 0.154 | 0.153 | 0.082 |
| OT37 | ATCACtGTGATGtTCGggAgAGG | 0 | 0 | 0 | 0.042 |
| OT38 | cTCtCgGTGgTGCTgGTCAAAGG | 0.045 | 0.061 | 0.069 | 0.082 |
| OT39 | gTgACAGTcATGCaCGTCcAAGG | 0.017 | 0.023 | 0.013 | 0.017 |
| OT40 | ATCACAcTGATtCcCtaCAAAGG | 0.051 | 0.097 | 0.024 | 0.077 |
| OT41 | ATgAgAGTGATttTCGTtAAAGG | 0.03 | 0.017 | 0 | 0.05 |
| OT42 | ATCACtGTGATGtTtacCAAAGG | 0.038 | 0.035 | 0.042 | 0.012 |
| OT43 | ATCACAGTGATGCTtccacAAGG | 0 | 0.02 | 0.034 | 0.012 |
| OT44 | gTaACAGTGgTGtTCGaCAAAGG | 0.113 | 0.209 | 0.142 | 0.192 |
| OT45 | ATCcCAaTcAgGCTCtTCAAAGG | 0.022 | 0.014 | 0.028 | 0.023 |
| OT46 | cTCACAcTGATGCaCtTCAtAGG | 0 | 0 | 0 | 0.01 |

| | | | | | |
|---|---|---|---|---|---|
| OT47 | AaCACAcTGAgGCTCtgCAAAGG | - | - | - | - |
| OT48 | ATggCAcTGATGCaCGaCAAAGG | 0.022 | 0.014 | 0.04 | 0.011 |
| OT49 | caCACtGTcATGtTCGTCAAAGG | 0.34 | 0.114 | 0.27 | 0.054 |
| OT50 | tTgACAGTGtTcCTaGTCAAAGG | 0.017 | 0.014 | 0.013 | 0 |
| OT51 | ATCAtAGgtATGtTgGTCAAAGG | 0 | 0.016 | 0.038 | 0.026 |
| OT52 | ATCACAcTGATGCcCtaCAtAGG | 0.011 | 0 | 0 | 0.021 |
| OT53 | ATCACAcTGATtCcCtgCAAAGG | 0.047 | 0.036 | 0.043 | 0.025 |
| OT54 | AaCAtAGcGtTGCTaGTCAAAGG | 0.049 | 0.043 | 0.087 | 0.119 |
| OT55 | ATCACAtgGATcCTCcTgAAAGG | 0.025 | 0 | 0 | 0 |
| OT56 | tTttCAaTGATGCTCaTCAAAGG | 0.023 | 0.015 | 0.018 | 0 |
| OT57 | tTCtCtGTcATGtTCGTCAAAGG | 0.027 | 0.052 | 0.02 | 0.019 |
| OT58 | ATCACAGTatTGgTCcaCAAAGG | 0.052 | 0.02 | 0.044 | 0.041 |
| OT59 | ATgctAGaGATGCTtGTCAAAGG | 0.029 | 0.01 | 0.017 | 0.078 |
| OT60 | ATCACAcTGATGCaCtaCAgAGG | 0 | 0 | 0 | 0.023 |
| OT61 | cTCACAcTGATGCaCtaCAAAGG | 0.051 | 0.052 | 0.061 | 0.018 |
| OT62 | tTgAtAGTGtTcCTCGTCAAAGG | - | - | - | - |
| OT63 | ATCACAGatATcaTgGTCAAAGG | 0.013 | 0 | 0.032 | 0.026 |
| OT64 | ATCttAGTcAaGCTaGTCAAAGG | - | - | - | - |
| OT65 | ATCAgAtTtATGCTCaTtAAAGG | - | - | - | - |
| OT66 | ATCtgAGTGATctTCGTCgAAGG | 0.033 | 0.02 | 0 | 0.027 |
| OT67 | ATggCAGTGtTcCTaGTCAAAGG | - | - | - | - |
| OT68 | ATCACAtTtATGCTtaTCtAAGG | 0.019 | 0.011 | 0.019 | 0.023 |
| OT69 | tcCACAGTGtTcCTaGTCAAAGG | 0.014 | 0.024 | 0.028 | 0.013 |
| OT70 | tTCttAGgGATGgTCGTCAAAGG | 0.042 | 0.02 | 0.024 | 0.013 |
| OT71 | AaCACAGTcATGCTCacCAgAGG | 3.006 | 2.67 | 2.831 | 0.935 |
| OT72 | AaaAgAGTGATGCTtaTCAAAGG | 0.018 | 0.012 | 0.018 | 0.029 |
| OT73 | cTtcCAGTGATGaTaGTCAAAGG | 0.051 | 0.021 | 0.02 | 0.043 |
| OT74 | ATCAaAGTGAgataCGTCAAAGG | 0.012 | 0.022 | 0 | 0.021 |
| OT75 | ATgAtAtTGAcGCTtGTCAAAGG | 0 | 0.055 | 0.02 | 0.053 |
| OT76 | ATCACgcTGATGggCcTCAAAGG | 0.012 | 0.016 | 0 | 0 |
| OT77 | ATagatGTGATGCTtGTCAAAGG | 0.012 | 0.02 | 0 | 0.022 |
| OT78 | gTCcCAtTGATGCaCGaCAAAGG | 0.017 | 0.046 | 0.051 | 0 |
| OT79 | tTgACAaTtATGCTCtTCAAAGG | 0.175 | 0.178 | 0.18 | 0.332 |
| OT80 | ATtAaAaTcATGtTCGTCAAAGG | 0.082 | 0.037 | 0.051 | 0.025 |
| OT81 | caCACAGTcATGtTCcTCAAAGG | 0 | 0.022 | 0.036 | 0.03 |
| OT82 | tTgACAaTcATGCTCtTCAAAGG | - | - | - | - |
| OT83 | tTCAtAGTGATGtTttTCAAAGG | 0.043 | 0.059 | 0.033 | 0.058 |
| OT84 | ATCACgcTcATGaTCcTCAAAGG | 0 | 0.03 | 0 | 0 |
| OT85 | ATCACAcTcATGgaCcTCAAAGG | 0 | 0.034 | 0.039 | 0.01 |
| OT86 | ATCAtAtTGAaGCcCtTCAAAGG | 0.027 | 0.053 | 0.079 | 0.053 |
| OT87 | ATCACAaTGATGgTCGgggAAGG | 0.268 | 0.358 | 0.301 | 0.273 |
| OT88 | ATCAtAaTGAaGCcCtTCAAAGG | 0.029 | 0.057 | 0.085 | 0.057 |
| OT89 | ATgAatGTtATGCTCtTCAAAGG | 0 | 0.038 | 0 | 0.052 |
| OT90 | ATCACAcTGATaCcCtaCAAAGG | 0.027 | 0.026 | 0.053 | 0.051 |
| OT91 | AatAtAaTGATtCTCGTCAAAGG | 0.022 | 0.02 | 0.013 | 0.036 |
| OT92 | ATgACtGTGtTcCTtGTCAAAGG | 0 | 0 | 0.122 | 0.074 |
| OT93 | cTCAaAGTcATGaTCtTCAAAGG | 0 | 0.026 | 0 | 0.022 |
| OT94 | cTCAatGaGATGCTCGaCAAAGG | 0.053 | 0.052 | 0.056 | 0.057 |
| OT95 | ATCACAcTtAaGCTCtTgAAAGG | 0.201 | 0.216 | 0.15 | 0.161 |
| OT96 | gTgACAGTGtTGCTtGTCgAAGG | 0.012 | 0.012 | 0.015 | 0 |
| OT97 | ATaACAacaATGaTCGTCAAAGG | 0.036 | 0.016 | 0.048 | 0.057 |
| OT98 | AaCACtGTGATGtTtGTCAgAGG | 0 | 0 | 0 | 0 |
| OT99 | ATCACgcTGATagTCcTCAAAGG | 0 | 0 | 0 | 0 |
| OT100 | gTgACAaTtATGCTCtTCAAAGG | 1.201 | 0.847 | 1.346 | 0.61 |

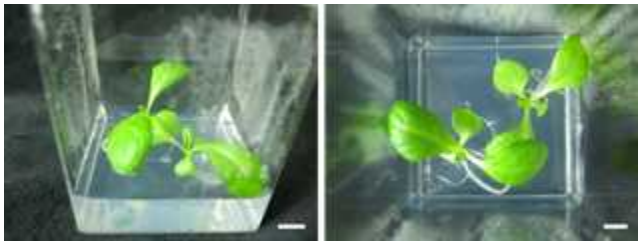
**Table 4. Indel frequencies at the on-target and 91 potential off-target sites in regenerated plantlets.** Indel frequencies were obtained with targeted deep sequencing. We couldn't analyze some potential off-target sites because of lack of reference sequence information or poor PCR condition.

## c. Germline transmission of mutant-alleles

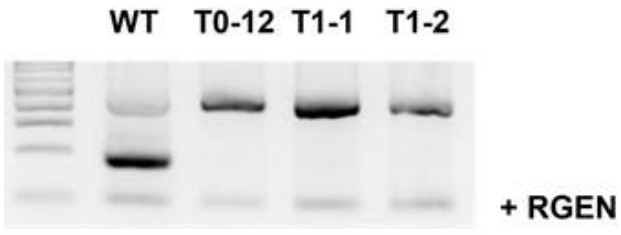
To test whether RGEN RNP-mediated genome modification is germline-transmissible, seeds were obtained from the T0-12 mutants and germinated. Two T1 plantlets were obtained and genomic DNA of these plantlets was isolated from a cotyledon (Figure 11a). To analyze the genotype of T1 plantlets, I performed RGEN RFLP assay and no cleavages of PCR amplicons were detected about the T0-12 mutant and two T1 plantlets, indicating that T1 plantlets were bi-allelic mutants like the T0-12 mutant (Figure 11b). To compare the sequence of mutant alleles, I carried out the sanger sequencing of PCR products which contained the BIN2 target site. The result obtained with T1 plantlets showed that 1-bp insertion was introduced at the cleavage site of BIN2-targeting RGEN and these plantlets were homozygous bi-allelic mutants. This result was in agreement with the result obtained with the T0-12 mutant, indicating that the mutant allele of T0-12 was stably transmitted to the next generation (Figure 11c). T1 plantlets were additionally obtained from T0-24 mutant, which contained heterozygous biallelic mutations. Genotypes of T1 plantlets were analyzed by targeted deep sequencing. I found that T1 plantlets derived from the T0-24 mutant were heterozygous bi-allelic mutants or homozygous bi-allelic mutants and frequencies of these mutants were in accordance with the law of Mendelian inheritance (Figure 12). In Summary, mutant plants could be obtained with RGEN RNP delivery in the first generations at a high frequency, and RGEN-mediated mutations were stably transmitted

into the next generation.

**a**



**b**




**c**

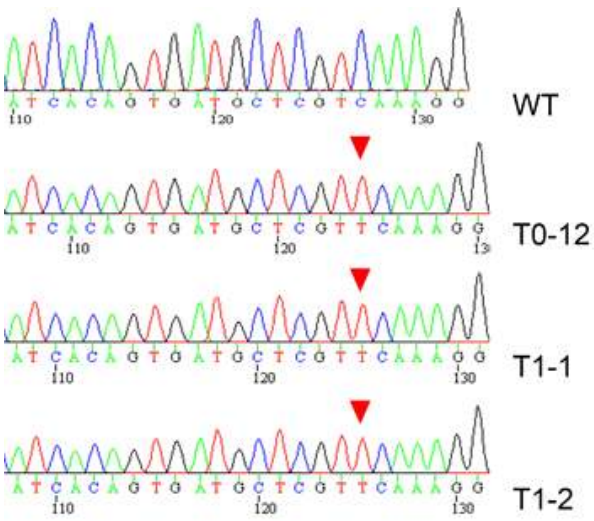


**Figure 11. Genotyping of T1 plantlets derived from T0-12 mutant.** (a) T1 plantlets derived from T0-12 mutant. Scale bars, 1 cm. (b) Genotyping of T1 plantlets by RGEN RFLP assay. T1 plantlets show bi-allelic mutant pattern. (c) Genotyping of T1 plantlets by Sanger sequencing. Red triangle indicates inserted nucleotide. WT, wild-type.

**a**

| T0 plant | | T1 plantlets | | | |
|---|---|---|---|---|---|
| Plant ID | mutation detected | Total No. of T1 plants | -1bp/-1bp | -1bp/+1bp | +1bp/+1bp |
| T0-24 | -1bp/+1bp | 18 | 4 (22.2 %) | 9 (50.0 %) | 5 (27.8 %) |

**b**

```
                ATCACAGTGATGCTCGT-CAAAGG   WT
T0-24           ATCACAGTGATGCTCG--CAAAGG   -1 (51.0 %)
                ATCACAGTGATGCTCGTtCAAAGG   +1 (48.9 %)
T1-24-1-A1      ATCACAGTGATGCTCGTtCAAAGG   +1 (50.8 %)
       -A2      ATCACAGTGATGCTCG--CAAAGG   -1 (49.0 %)
T1-24-2         ATCACAGTGATGCTCG--CAAAGG   -1 (99.7 %)
T1-24-3-A1      ATCACAGTGATGCTCGTtCAAAGG   +1 (50.4 %)
       -A2      ATCACAGTGATGCTCG--CAAAGG   -1 (49.3 %)
T1-24-4         ATCACAGTGATGCTCGTtCAAAGG   +1 (99.8 %)
T1-24-5         ATCACAGTGATGCTCGTtCAAAGG   +1 (99.7 %)
T1-24-6-A1      ATCACAGTGATGCTCGTtCAAAGG   +1 (51.2 %)
       -A2      ATCACAGTGATGCTCG--CAAAGG   -1 (48.5 %)
T1-24-7-A1      ATCACAGTGATGCTCG--CAAAGG   -1 (51.5 %)
       -A2      ATCACAGTGATGCTCGTtCAAAGG   +1 (48.3 %)
T1-24-8         ATCACAGTGATGCTCG--CAAAGG   -1 (99.8 %)
T1-24-9         ATCACAGTGATGCTCG--CAAAGG   -1 (99.7 %)
T1-24-10-A1     ATCACAGTGATGCTCG--CAAAGG   -1 (50.8 %)
        -A2     ATCACAGTGATGCTCGTtCAAAGG   +1 (49.1 %)
T1-24-11        ATCACAGTGATGCTCG--CAAAGG   -1 (99.7 %)
T1-24-12        ATCACAGTGATGCTCGTtCAAAGG   +1 (99.6 %)
T1-24-13-A1     ATCACAGTGATGCTCG--CAAAGG   -1 (50.9 %)
        -A2     ATCACAGTGATGCTCGTtCAAAGG   +1 (48.9 %)
T1-24-14-A1     ATCACAGTGATGCTCG--CAAAGG   -1 (50.1 %)
        -A2     ATCACAGTGATGCTCGTtCAAAGG   +1 (49.8 %)
T1-25-15-A1     ATCACAGTGATGCTCGTtCAAAGG   +1 (50.5 %)
        -A2     ATCACAGTGATGCTCG--CAAAGG   -1 (49.3 %)
T1-24-16-A1     ATCACAGTGATGCTCG--CAAAGG   -1 (50.2 %)
        -A2     ATCACAGTGATGCTCGTtCAAAGG   +1 (49.6 %)
T1-24-17        ATCACAGTGATGCTCGTtCAAAGG   +1 (99.6 %)
T1-24-18        ATCACAGTGATGCTCGTtCAAAGG   +1 (99.8 %)
```

**Figure 12. Genotyping of T1 plantlets derived from T0-24 mutant.** (a) Summary table for genotyping analysis of T1 plantlets obtained from T0-24 mutant. (b) Mutant sequences of T1 plantlets. Targeted deep sequencing was performed and percentages of mutant reads were indicated in parentheses. PAM sequence is shown in bold and inserted nucleotide is shown in lower case. Guide sequence is underlined.

# Discussion

In this study, I designed a new platform for plant genome editing via preassembled CRISPR-Cas9 ribonucleoprotiens. RGEN RNPs induced targeted genome modifications in plant protoplasts derived from Arabidopsis thaliana, tobacco, rice, and lettuce efficiently with frequencies that ranged from 8.4% to 44% and whole plants harboring the gene-edited allele were successfully regenerated from RGEN RNP-transfected lettuce protoplasts at a frequency of 46% in the first generation. No apparent off-target effects were detected in the regenerated plantlets and I found that gene-edited plants transmitted the mutant alleles to the next generation in accordance with the law of Mendelian inheritance.

Now, programmable nucleases including ZFNs, TALENs, and RGENs have become an indispensable tool for targeted genome engineering in plants (Li et al., 2013; Li et al., 2012; Nekrasov et al., 2013; Shan et al., 2013; Zhang et al., 2010; Zhang et al., 2013). In general, delivery of programmable nucleases in plants have been done by transfecting plasmids which encode an expression cassette of nucleases or infecting plants with pathogens such as Agrobacterium that can transfer a genetic element called T-DNA into the host genome. These methods, both depend upon plant cells for transcription and translation of these nucleases, so cell-type specific promoters and codon optimization are required for the efficient genome editing with programmable nucleases in plant cells, demanding a laborious work for

each plant species. However, these efforts are unnecessary for RGEN RNPs because translation and transcription of nucleases have already been completed in vitro. RGEN RNPs cleave the target site immediately after transfection and induce targeted genome modification at early time-point (Kim et al., 2014b), reducing a likelihood of obtaining a mosaic phenotype in regenerated plantlets. RGEN RNPs are also rapidly degraded by endogenous proteinase in cells, which might reduce the off-target effects induced by RGENs (Kim et al., 2014b). In this study, I found that targeted mutagenesis was achieved in regenerated plantlets at a high frequency of 46% and most of genome-edited plantlets contained bi-allielic mutations at the target site (14 of 16 mutants), indicating that delivery of RGEN RNP might be an effective tool for multiplex genome editing in plants.

One difficulty with generating a genome-edited plant with delivery of RGEN RNP is the optimization process for regenerating whole plants from protoplasts. In this study, I used cell-wall removed protoplasts as a starting material for PEG-mediated transfection. Protoplasts are generally suitable for delivery of biomolecules into plant cells, but are fragile because the rigidity of plant cells is highly sensitive to turgor pressure in case of cell wall-removed condition. Further studies are needed to test whether other sources such as callus can be used as a starting material for RGEN RNP-mediated genome editing.

Improving the RGEN RNP-mediated genome editing system more efficient in plant cells is another important issue, because

handling a lot of callus derived from RGEN RNP-transfected cells is laborious and time-consuming. Recently, several groups reported that RGEN RNP-mediated genome editing with a chemically modified sgRNA could enhance the efficiency of targeted genome editing in human cells (Hendel et al., 2015; Rahdar et al., 2015). These studies showed that introducing chemical modifications such as 2'-O-mehtyl-3' phosphorothioate and 2'-O-methyl 3' thioPACE at the both termini of sgRNA yielded and enhanced genome editing efficiency at the target site. Plant genome editing with this method might be a good option for increasing the on-target indel rate. Chemical modification of RNAs also can be combined with currently characterized CRISPR-Cpf1 system (Zetsche et al., 2015a). CRISPR-Cpf1 system consists of Cpf1 nucleases and crRNAs. The length of crRNA is 43-nt, much shorter than that of sgRNA. This feature is advantageous for synthesizing and modifying the RNAs.

In this study, I selected the BRASSINOSTEROID INSENSITIVE 2 (BIN2) gene for targeted mutagenesis in lettuce. In fact, this gene has been known having different phenotypes according to a type of mutation. BIN2-KO mutant showed enhanced tolerance to various abiotic stresses in rice (Koh et al., 2007), whereas missense mutation yielded a semi-dominant dwarf phenotype in Arabidopsis (Choe et al., 2002). I only generated BIN2 knock-out mutants by delivery of RGEN RNPs, but further studies are needed to develop a system enabling the genetic modification such as point mutation combined with the RNP delivery method for modeling of all the

genetic alterations found in plant systems. Recently, several groups independently reported the base editing systems enabling a single nucleotide editing without DNA DSBs at the target site with Cas9-deaminase fusion proteins (Hess et al., 2016; Komor et al., 2016; Ma et al., 2016; Nishida et al., 2016). DNA-free base editing that combines RNPs delivery with base editor might broaden the field of CRISPR-based plant genome editing in the near future

In summary, RGEN RNP delivery enables DNA-free genome editing in plant cells efficiently. Gene-edited plants are successfully regenerated from RGEN RNP-transfected protoplasts and mutant DNA sequences are stably transmitted to the next generation. I propose that this method might be an alternative for current GMO regulations and facilitate the usage of this tool in plant biotechnology and agriculture.

# Chapter 2. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells

# Introduction

The CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) system is an adaptive immune response in bacteria and archaea, which acts against foreign genetic elements (Barrangou et al., 2007; Horvath and Barrangou, 2010; Sorek et al., 2008; Sorek et al., 2013; Terns and Terns, 2011). In this system, CRISPR RNAs (crRNAs) guide cas effector proteins to the invading genetic elements in a sequence-specific manner, which yields silencing of foreign elements (Horvath and Barrangou, 2010). CRISPR-Cas systems consist of the three steps for the defense activity. First, short fragment of foreign DNA (protospacers) is integrated into the host genome at the end of the CRISPR array as known as adaptation process. Next, pre-crRNAs are transcribed from the CRISPR array and pre-crRNA is subjected to maturation, yielding processed crRNAs. Finally, crRNAs interact with a single Cas effector molecule or Cas protein complex, guide ribonuleoprotein complexes to the target substrates, and cleave target nucleic acids (Horvath and Barrangou, 2010).

CRISPR-Cas systems can be classified into two groups based on the format of Cas effector (Makarova and Koonin, 2015). Class 1 systems which include CRISPR-Cas type I system (Hochstrasser et al., 2014; Mulepati et al., 2014; Wiedenheft et al., 2011) and CRISPR-Cas type III system (Hale et al., 2014; Staals et al., 2014) harness crRNAs and Cas protein complexes composed of multi-components. On the other

hand, class 2 system, including CRISPR-Cas type II utilizes a processed crRNA and single effector protein such as the cas9 protein (Jinek et al., 2012). Of the CRISPR-Cas type II systems, CRISPR-Cas9 system derived from the pathogen Streptococcus pyogenes has been widely utilized for targeted genome engineering in human cells, mouse, model organisms, and plants (Cho et al., 2013a; Cho et al., 2013b; Cong et al., 2013; Hwang et al., 2013; Jiang et al., 2013; Jinek et al., 2013; Li et al., 2013; Mali et al., 2013; Nekrasov et al., 2013; Shan et al., 2013; Sung et al., 2014; Wang et al., 2013). Cas9 protein induces DNA double-strand breaks (DSBs) at the chromosomal target site guided by gRNAs. This cleavage produces blunt DSBs that are repaired by endogenous repair mechanisms such as non-homologous end joining (NHEJ) and homologous recombination (HR) (Kim and Kim, 2014). Especially, error-prone NHEJ repair mechanism is predominant in eukaryotic cells (Shrivastav et al., 2008), yielding indels as a major pattern of RNA-guided genome editing.

In a recent study, a new type of CRISPR Class 2 system has been identified in the bacterial genome, assigned to a type V system (Makarova et al., 2015). This system has different characteristics compared with type II CRISPR-Cas9 system (Zetsche et al., 2015a). First, this system contains a protein called Cpf1 (CRISPR from Prevotella and Francisella 1) and Cpf1 act as a single effector molecule instead of Cas9 protein. Second, pre-crRNAs are processed into the mature crRNAs without trans-activating crRNA (tracrRNA) (Deltcheva et al., 2011). Third, Cpf1-crRNA complexes recognize thymidine-rich DNA

sequence as a protospacer-adjacent motif (PAM) unlike Cas9 protein which recognizes the guanine-rich DNA sequence. Finally, CRISPR-Cpf1 system cleaves target DNA, yielding staggered ends at the cleavage site unlike Cas9 protein, which produces blunt ends at the target site. 16 Cpf1-family proteins were tested for targeted genome engineering in human cells, two of these derived from Acidaminococcus sp. BV3L6 and Lachnospiraceae bacterium ND2006 showed a robust genome editing in HEK293T cells (Zetsche et al., 2015a). This result indicates that targetable site by CRISPR-endonucleases increases. Moreover, CRISPR-Cpf1 systems might facilitate a precise knock-in by directing the orientation of transgene fragment.

Despite these benefits, the application of the CRISPR-Cpf1 system in genome-editing has been hampered by low on-target indel rates (Zetsche et al., 2015a) and a lack of studies evaluating a target specificity of Cpf1 nucleases. Especially, specificities of Cpf1 endonucleases are a very important issue. Target specificity of Cpf1 endonucleases is determined by 23-nucleotide length of guide sequence and there are hundreds of homologous sites that differ from the on-target sequence by up to 5-nt mismatches in the human genome. In case of CRISPR-Cas9, target specificity of RGEN has been studied by targeted analysis of predicted off-target sites based on sequence homology (Cho et al., 2014; Fu et al., 2013; Hsu et al., 2013) or unbiased methods (Frock et al., 2015; Kim et al., 2015; Tsai et al., 2015). These studies found that RGEN tolerates mismatches between guide sequence of sgRNA and target DNA and theses effects are

dependent on the number of mismatches, distribution of mismatches, and the sequence context. In some case, DNA/RNA bulge obtained by sgRNA-Cas9 protein interaction caused off-target effects in the human genome (Kim et al., 2015). Based on these findings, several methods were suggested for reducing off-target effects (Cho et al., 2014; Fu et al., 2014; Guilinger et al., 2014; Kim et al., 2014b; Kleinstiver et al., 2016; Slaymaker et al., 2016; Tsai et al., 2014). To utilize the potential of Cpf1, characterization of target specificities of CRISPR-Cpf1 endonucleases in the human genome context might be necessary like SpCas9.

Here, I optimize the Cpf1-mediated genome editing system in human cells and analyze the target specificity of Cpf1 endonucleases in the context of the human genome (Kim et al., 2016a). Cpf1 endonucleases induce targeted genome modifications in human cells at a high frequency and are highly sensitive to mismatches in the 3' PAM proximal region. These results suggest that CRISPR-Cpf1 endonuclease is suitable for precise genome engineering.

# Materials and Methods

## 1. Cas9 and Cpf1 ribonucleoproteins

Recombinant Cas9 protein was purchased from ToolGen. SpCas9 sgRNAs and Cpf1 crRNAs were synthesized by in vitro transcription using T7 RNA polymerase as described previously. For preparation of recombinant Cpf1 proteins, the protein sequences were codon optimized for E.coli expression. Synthetic oligonucleotide fragments were cloned into bacterial protein expression vectors that encode 6xHis and maltose-binding protein tags (pMAL-c5x, New England BioLabs; and pDEST-hisMBP). The resulting expression vectors were used to transform Rosetta expression cells (EMD Millipore). 2 l of Luria broth (LB) growth medium with 50 mg/ml carbenicilin were inoculated with 10 ml overnight culture of Rosetta cells containing Cpf1 plasmids. Cells were cultured at 37 °C until OD600 reached 0.6, and then cooled to 16 °C before induction with 0.5 mM IPTG for 14-18 h. Cells were then harvested and frozen at − 80 °C until protein purification. For protein purification, the cell pellet was lysed by sonication in 50 ml of lysis buffer (50 mM, HEPES pH 7, 200 mM NaCl, 5 mM MgCl2, 1 mM DTT, 10 mM imidazole) supplemented with lysozyme (Sigma) and protease inhibitor (Roche complete, EDTA-free). The cell lysate was cleared by centrifugation at 16,000g for 30 min, followed by passage through a syringe filter (0.22 μm). The cleared lysate was applied to a nickel column (Ni-NTA

agarose, Qiagen), washed with 2 M salt, and eluted with 250 mM imidazole. The eluted protein solution was buffer exchanged and concentrated with lysis buffer without magnesium and imidazole. Purified Cpf1 proteins were examined with SDS-PAGE. Cpf1 protein activity was assessed by an in vitro cleavage assay.

## 2. Plasmids encoding Cpf1 and crRNAs

Cpf1-encoding vectors were obtained from Addgene. The Cpf1 expression cassette was cloned into pcDNA3.1 backbone vector. Amino acid sequence of Cpf1 was human-codon optimized and CMV promoter was used for strong expression in human cells. The nuclear localization signal of nucleoplasmin and HA epitope were tagged at the C terminus of Cpf1 gene. crRNA-encoding vector was modified from pRG2-EXT vector which encodes sgRNA of SpCas9. crRNA vector was designed to human U6 promoter-driven transcription and poly-T signal was used for termination of crRNA transcription. Two BsmBI sites were inserted between 5' repeat sequence of crRNA and poly-T signal for replacing guide sequence of crRNA.

## 3. Cell culture and transfection conditions

HEK293T/17 (ATCC, CRL-11268) cells were maintained with Dulbecco's Modified Eagle's Media (Welgene) supplemented with 10% FBS (Welgene) and 1% Antibiotics (Welgene). Cells were subcultured every two to three days. For Cpf1-mediated gene editing, 1.5 × 105

cells were seeded into 24-well-plates before transfection. Next day, Cpf1 expression vector (500 ng) and crRNA-encoding vector (500 ng) were transfected into HEK293T17 cells using lipofectamine 2000 (Invitrogen). Condition for liopfection was based on manufacturer's instructions. After 72 h post-transfection, cells were harvested and genomic DNA was purified with DNeasy Blood & Tissue Kit (Qiagen).

## 4. Isolation of mutant clones

For isolation of cpf1-mediated mutant clones, Hela (ATCC, CCL-2) cells was used. 8 × 104 cells were seeded into 24-well-plates before transfection and lipofection was performed using a protocol for HEK293T/17 cell transfection. Transfected cells were diluted and diluted cells were seeded into 96-well-plates at 0.3 cells/well condition. Single cell-derived clones were analyzed after two to three weeks.

# Results

## 1. Genome editing in human cells using Cpf1 system

### a. Optimization of Cpf1 system

In the previous study, Cpf1 system showed a high success rate of gene modification at the randomly selected sites, but the efficiencies of Cpf1 nucleases were generally less than 10 %. (Zetsche et al., 2015a). Before investigating a specificity of Cpf1 system, I optimized the condition for genome editing with Cpf1 nucleases in human cells. First, I performed the cell test with Cpf1 expression plasmids and crRNA-encoding amplicons as previously described. HEK293T/17 cells were seeded at 24-well plates and Cpf1 expression plasmids and crRNA-encoding amplicons were transfected into cells using lipofectamine 2000 next day. Several ratios of Cpf1 plasmid and crRNA-encoding amplicon were tested. Genomic DNA was isolated 72 h post-transfection. Gene modification efficiencies were measured by T7E1 analysis. Both LbCpf1 and AsCpf1 showed on-target indel rates similar to those previously reported but, gene modification efficiencies were still low (Figure 13).

To improve an efficiency of Cpf1 system, I designed a crRNA-encoding plasmid (Figure 14). Theoretically, circular plasmid DNA is more stable than PCR amplicon, as there are no ends from

which degradation can easily occur by exonucleases in human cells. Moreover, PCR amplicon can induce transcription of incorrect RNA due to synthesis-failed oligonucleotide templates, potentially causing off-target DNA cleavages at sites harboring an RNA bulge (Kim et al., 2016b). To verify an efficiency of Cpf1-mediated genome editing with crRNA-encoding plasmid, I transfected AsCpf1 expression plasmid and crRNA plasmid targeting the DNMT1-3 site. I measured the indel rates at the DNMT1-3 site by T7E1 assay and targeted deep sequencing and found that the highly efficient genome editing occurred at the DNMT1-3 site (Figure 15). To compare the efficiencies of Cpf1-mediated genome editing in HEK293T/17 cells transfected with either plasmids or PCR amplicons encoding crRNA directly, I chose three endogenous sites and performed the cell test. I found that the frequency of targeted mutagenesis increased by 2-30-fold at three endogenous target sites using crRNA plasmids rather than amplicons (Figure 16). To optimize genome editing efficiency and to avoid unwanted off-target effects at such sites, I used plasmids encoding crRNAs throughout the study.

Next, I examined whether other Cpf1-family protein could be used for efficient genome editing in human cells with the optimized Cpf1 system. I additionally designed a crRNA-encoding plasmid for FnCpf1 and MbCpf1-mediated genome editing (Figure 17). FnCpf1 and MbCpf1 have been shown to be inefficient or inactive in human cells (Zetsche et al., 2015a), but I found that FnCpf1 and MbCpf1 could induce efficient genome editing at the endogenous sites with Cpf1

expression plasmids and crRNA-encoding plasmids. I further analyzed crRNA orthogonality with four Cpf1 orthologs from a Lachnospiraceae bacterium (LbCpf1), Acidaminococcus sp. (AsCpf1), Francisella novicida (FnCpf1) and Moraxella bovoculi 237 (MbCpf1). Cpf1 orthologs were co-transfected into HEK293T/17 cells with plasmids encoding crRNA orthologs in various combinations. Each Cpf1 ortholog was most efficient with its cognate crRNA. However, all four Cpf1 orthologs in combination with non-orthologonal crRNAs cleaved chromosomal target sites (Figure 17d). This result suggests that the Cpf1 protein is less sensitive to the difference in 5' loop of crRNA and further engineering is necessary for orthogonal genome manipulation with four Cpf1 orthologs.

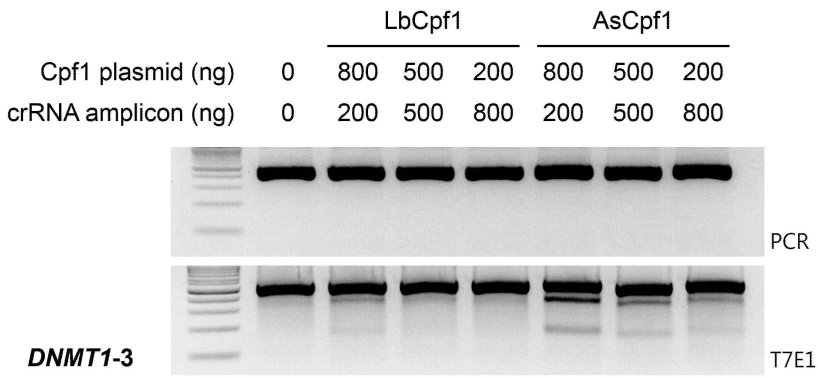



**Figure 13. Genome editing with Cpf1 plasmids and crRNA-encoding amplicon.** Cpf1-mediated gene editing in HEK293T17 cells. Cpf1 expression plasmid and crRNA-encoding PCR amplicon were transfected into cells using lipofectamine 2000 and genomic DNA was isolated 72 h post-transfection. Targeted gene editing events were detected by the T7E1 assay.

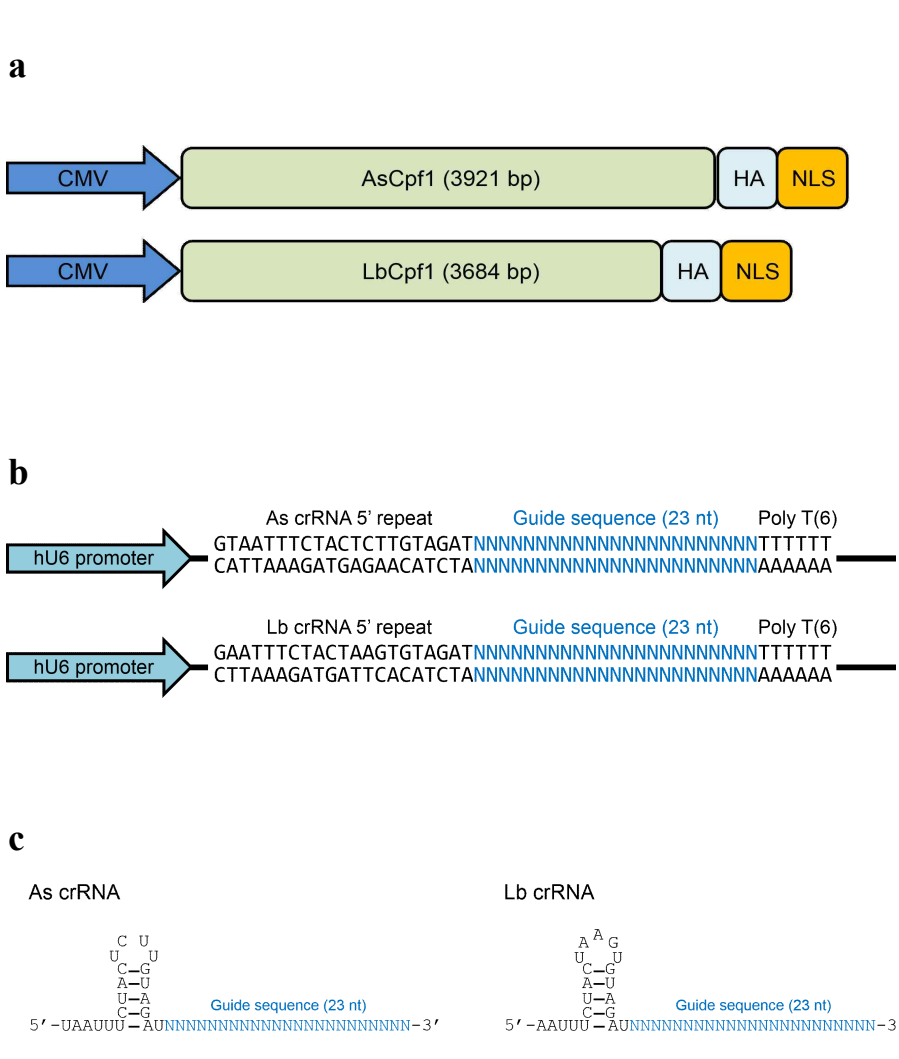
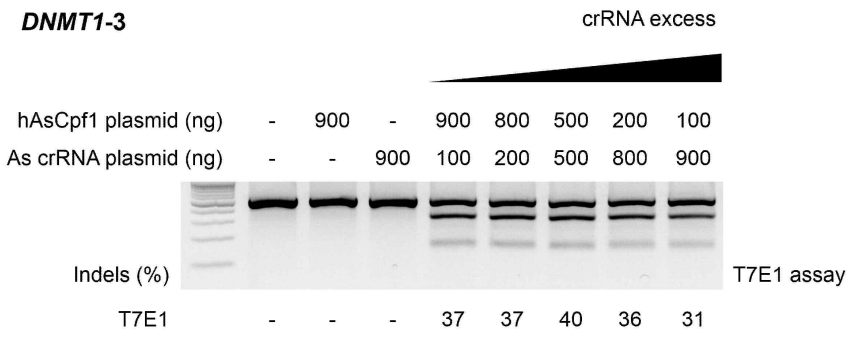



**Figure 14. DNA constructs for Cpf1-meidated genome editing.** (a) Plasmid DNA constructs for Cpf1 expression. (b) Plasmid DNA constructs for hU6 promoter-driven crRNA transcription. (c) Secondary structure of crRNAs.

**a**




**b**

| | Mock | Cpf1 only | crRNA only | Cpf1 900 crRNA 100 | Cpf1 800 crRNA 200 | Cpf1 500 crRNA 500 | Cpf1 200 crRNA 800 | Cpf1 100 crRNA 900 |
|---|---|---|---|---|---|---|---|---|
| Indel count | 0 | 0 | 0 | 2722 | 3942 | 3442 | 3795 | 2492 |
| Total count | 9950 | 9127 | 12644 | 6777 | 9042 | 7456 | 8652 | 6735 |
| Indel (%) | 0.0 | 0.0 | 0.0 | 40 | 44 | 46 | 44 | 37 |

**Figure 15. Condition optimization of Cpf1-mediated genome editing.** (a) Several dose conditions of Cpf1 expression plasmids and crRNA-encoding plasmids were tested. Targeted gene editing events were analyzed by the T7E1 assay. (b) Raw indel counts obtained from targeted deep sequencing analysis.

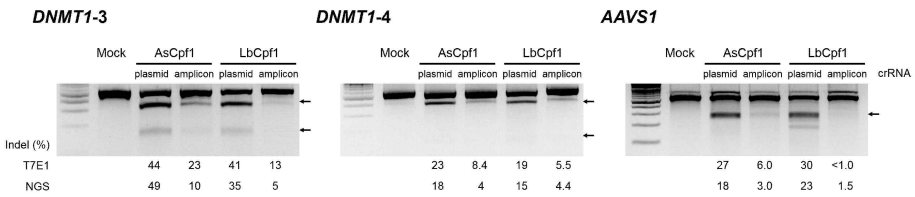
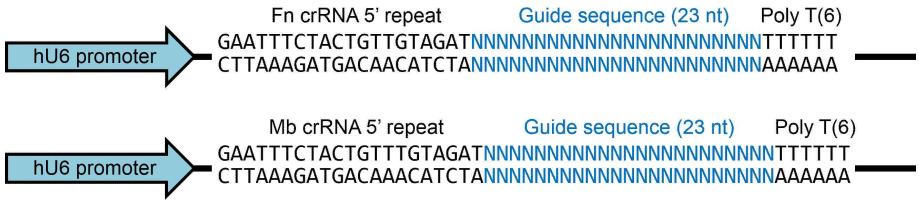



**Figure 16. Comparison of crRNA delivery methods.** crRNA-encoding plasmids or amplicons were used for Cpf1-mediated gene editing in three endogenous sites. Indel rates were analyzed by the T7E1 assay and targeted deep sequencing. Arrows indicate cleaved DNA bands.

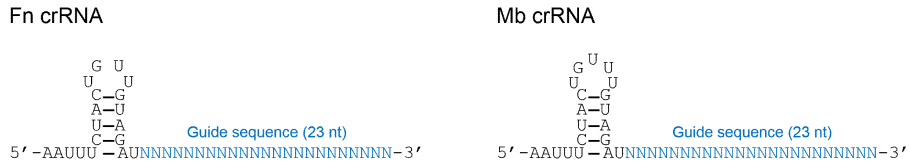
**a**


CMV
FnCpf1 (3900 bp)
HA
NLS
CMV
MbCpf1 (4119 bp)
HA
NLS

**b**


Fn crRNA 5' repeat
Guide sequence (23 nt)
Poly T(6)
hU6 promoter
GAATTTCTACTGTTGTAGATNNNNNNNNNNNNNNNNNNNNTTTTTT
CTTAAAGATGACAACATCTANNNNNNNNNNNNNNNNNNNNAAAAAA
Mb crRNA 5' repeat
Guide sequence (23 nt)
Poly T(6)
hU6 promoter
GAATTTCTACTGTTTGTAGATNNNNNNNNNNNNNNNNNNNNTTTTTT
CTTAAAGATGACAAACATCTANNNNNNNNNNNNNNNNNNNNAAAAAA

**c**


Fn crRNA
Guide sequence (23 nt)
5'-AAUUU-AUNNNNNNNNNNNNNNNNNNNN-3'
Mb crRNA
Guide sequence (23 nt)
5'-AAUUU-AUNNNNNNNNNNNNNNNNNNNN-3'

**d**

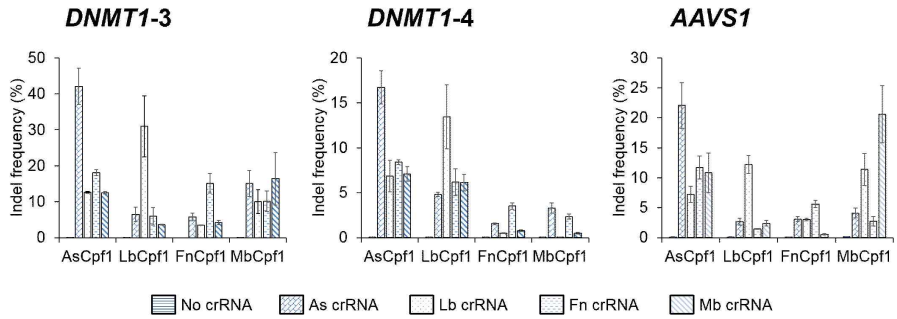

DNMT1-3
DNMT1-4
AAVS1
Indel frequency (%)
AsCpf1 LbCpf1 FnCpf1 MbCpf1
No crRNA
As crRNA
Lb crRNA
Fn crRNA
Mb crRNA

**Figure 17. Genome editing using plasmids encoding Cpf1 orthologs and crRNAs.** (a) Plasmid DNA constructs for FnCpf1 and MbCpf1 expression. (b) Plasmid DNA constructs for hU6 promoter-driven Fn crRNA and Mb crRNA transcription. (c) Secondary structures of Fn crRNA and Mb crRNA. (d) crRNA orthogonality with four Cpf1 orthologs. Plasmids encoding Cpf1 orthologs were co-transfected with plasmids encoding crRNAs, in various combinations, into HEK293T17 cells. Indel rates were measured by targeted deep sequencing. Error bars indicate s.e.m.

## b. Isolation of mutant clones

To investigate whether Cpf1 system could be used for obtaining mutant clones, I transfected the DNMT1-3 targeting crRNA plasmids and Cpf1-encoding plasmids into Hela cells and carried out limiting dilution of transfected cells 72 h post-transfection and diluted cells were seeded at 96-well-plates. After 3 weeks, single cell-derived clones were genotyped by targeted deep sequencing. 6 of 13 clones for AsCpf1 and 4 of 28 clones for LbCpf1 contained mutant-alleles at the target sites (Figure 18), demonstrating that Cpf1-induced mutations were maintained stably and Cpf1 system is a powerful tool for generating mutant clones.

**a**

*DNMT1*-3 mutated clone (AsCpf1) : 46 % (6/13)

```
     ATCAGTACGTTAATGTTTCCTGATGGTCCATGTCTGTTACTCGCCTGTCAAGTGGCG   WT
#1   ATCAGTACGTTAATGTTTCCTGATGGTCAA----------------------GTGGCG  -21 (60.8 %)
     ATCAGTACGTTAATGTTTCCTGATGGTCCATGTCTGTTACTCGCCTGTCAAGTGGCG   WT (39.1 %)
#3   ATCAGTACGTTAATGTTTCCTGATGG-----------------------------CG  -29 (57.6 %)
     ATCAGTACGTTAATGTTTCCTGATGGTCCATG-----------CCTGTCACAAGGCG  -11 (42.3 %)
#7   ATCAGTACGTTAATGTTTCCTGATGGTCCATG--------TCGCCTGTCAAGTGGCG  -8  (99.4 %)
#10  ATCAGTACGTTAATGTTTCCTG-------------------------TCAAGTGGCG  -25 (44.5 %)
     ATCAGTACGTTAATGTTTCCTGATGGTCCA---------CTCGCCTGTCAAGTGGCG  -9  (28.6 %)
     ATCAGTACGTTAATGTTTCCTGATGGTCCAT--------CTCGCCTGTCAAGTGGCG  -8  (26.7 %)
#11  ATCAGTACGTTAATGTTTCCTGATGGTCCA---------------TGTCAAGTGGCG  -15 (41.1 %)
     ATCAGTACGTTAATGTTTCCTGATGGTCCATGT------CTCGCCTGTCAAGTGGCG  -6  (36.5 %)
     ATCAGTACGTTAATGTTTCCTGATGGTCCATGTCT--TACTCGCCTGTCAAGTGGCG  -2  (22.1 %)
#12  ATCAGTACGTTAATGTTTCCTGATGGTCCA-------TACTCGCCTGTCAAGTGGCG  -7  (34.6 %)
     ATCAGTACGTTAATGTTTCCTGATGGTCCAT--------CTCGCCTGTCAAGTGGCG  -8  (33.9 %)
     ATCAGTACGTTAATGTTTCCTGATGGTCCATGTCTGTTACTCGCCTGTCAAGTGGCG   WT (31.5 %)
```

**b**

*DNMT1*-3 mutated clone (LbCpf1) : 14 % (4/28)

```
     ATCAGTACGTTAATGTTTCCTGATGGTCCATGTCTGTTACTCGCCTGTCAAGTGGCG   WT
#4   ATCAGTACGTTAATGTTTCCTGATGGTCCATGT---TTACTCGCCTGTCAAGTGGCG  -3  (99.8 %)
#11  ATCAGTACGTTAATGTTTCCTGATGGTCCA---------------TGTCAAGTGGCG  -15 (99.8 %)
#16  ATCAGTACGTTAATGTTTCCTGATGGTCCATGTCTGTTACTCGCCTGTCAAGTGGCG   WT (63.9 %)
     ATCAGTACGTTAATGTTTCCTGATGGTCCA------TTACTCGCCTGTCAAGTGGCG  -6  (36.1 %)
#18  ATCAGTACGTTAATGTTTCCTGATGGT------------------------------  -43 (99.6 %)
```

**Figure 18. Isolation of Cpf1-mediated mutant clones.** (a) DNA sequences of DNMT1-3 mutant clones obtained with AsCpf1. Hela cells were transfected with AsCpf1 expression plasmid and crRNA-encoding plasmid. Transfected cells were diluted and diluted cells were seeded at 96-well plates. Genotypes of colonies were analyzed by targeted deep sequencing. PAM sequence is shown in bold and guide sequence is underlined. WT, wild-type. (b) DNA sequences of DNMT1-3 mutant clones obtained with LbCpf1.

## c. Comparison of the mutation frequencies obtained with Cpf1 and SpCas9

To compare the frequency of targeted mutations by AsCpf1, LbCpf1 and SpCas9, I chose ten chromosomal target sites which contained two PAM sequences, one recognized by Cpf1 (5'-TTTN-3') and the other recognized by SpCas9 (5'-NGG-3'). Each RGEN had a wide range of mutation frequencies at these sites in HEK293T/17 cells (Table 5). SpCas9 was more efficient than Cpf1, with an average mutation rate of 32 ± 4%. LbCpf1 and AsCpf1 had mutation frequencies of 19 ± 6% and 20 ± 5% (Figure 19 and Table 5), respectively, which is about the same frequency at Staphylococcus aureus Cas9 (Ran et al., 2015).

| Target ID | Gene | Chromosome | Position | Target site | Indel frequency (%) | | |
|---|---|---|---|---|---|---|---|
| | | | | | LbCpf1 | AsCpf1 | SpCas9 |
| *DNMT1*-3 | *DNMT1* (intron) | Chr 19 | 10133739 | taatg**TTTC**CTGATGGTCCATGTCTGTTACTCgcctgtcaagtggcgtg | 31 ± 3.5 | 42 ± 8.1 | 51 ± 7.2 |
| *DNMT1*-4 | *DNMT1* | Chr 19 | 10133686 | ttcct**TTTA**TTTCCCTTCAGCTAAAATAAAGGaggaggaagctgctaag | 6.4 ± 2.6 | 9.7 ± 1.5 | 12 ± 1.5 |
| *EMX1*-2 | *EMX1* | Chr 2 | 72933764 | tgtac**TTTG**TCCTCCGGTTCTGGAACCACACCttcacctgggccaggga | 18 ± 5.8 | 13 ± 3.0 | 38 ± 7.8 |
| *AAVS1* | Intergenic | Chr 19 | 55115548 | taagg**TTTG**CTTACGATGGAGCCAGAGAGG**G**ATcctgggagggagagctt | 19 ± 3.5 | 23 ± 4.2 | 36 ± 6.9 |
| *CCR5*-1 | *CCR5* | Chr 3 | 46373006 | tttgg**TTTT**GTGGGCAACATGCTGGTCATCCTcatcctgataaactgca | 13 ± 1.5 | 10 ± 1.3 | 23 ± 3.5 |
| *CCR5*-9 | *CCR5* | Chr 3 | 46373664 | aattc**TTTG**GCCTGAATAATTGCAGTAGCTCTaacaggttggaccaagc | 18 ± 3.6 | 21 ± 5.4 | 20 ± 4.9 |
| *HPRT1*-1 | *HPRT1* | Chr X | 134475241 | aattc**TTTG**CTGACCTGCTGGATTACATCAAAgcactgaatagaaatag | 20 ± 1.4 | 20 ± 2.3 | 36 ± 2.4 |
| *HPRT1*-4 | *HPRT1* | Chr X | 134486436 | ttact**TTTA**TGTCCCCTGTTGACTGGTCATTCtagttaaaaaaaaaaaa | 65 ± 1.7 | 48 ± 4.5 | 43 ± 2.3 |
| *VEGFA*-2 | *VEGFA* | Chr 6 | 43770839 | gaaac**TTTT**CGTCCAACTTCTGGGCTGTTCTCgcttcggaggagccgtg | 1.3 ± 0.20 | 5.4 ± 0.81 | 43 ± 6.0 |
| *HBB* | *HBB* | Chr 11 | 5226711 | ggttc**TTTG**AGTCCTTTGGGGATCTGTCCACTcctgatgctgttatggg | 2.3 ± 0.24 | 3.3 ± 0.53 | 19 ± 3.8 |

**Table 5. On-target activities of AsCpf1, LbCpf1 and SpCas9.** Summary of on-target activities of AsCpf1, LbCpf1 and SpCas9 at 10 endogenous sites. Cpf1 PAM sequence is shown in bold and Cpf1 target sequence is shown in upper case. Cas9 PAM sequence is underlined.

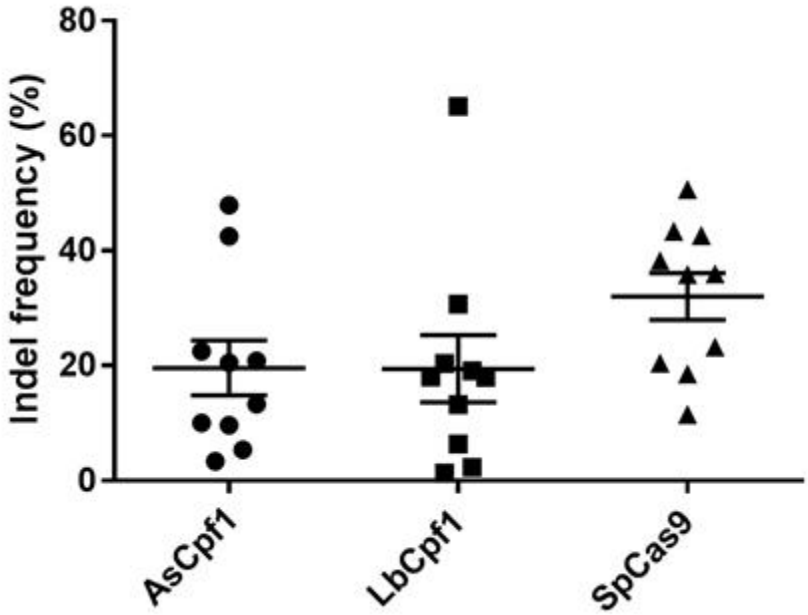


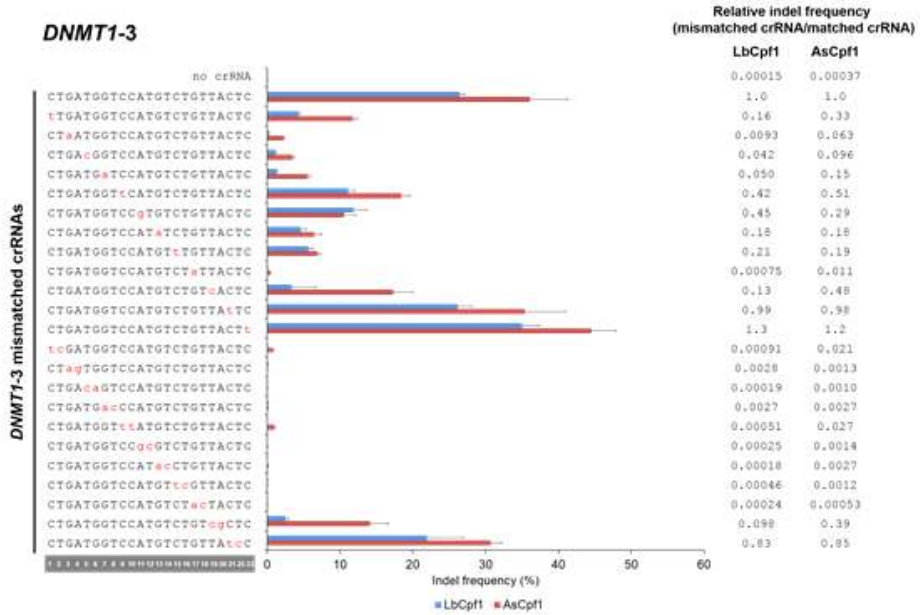

**Figure 19. On-target indel rates obtained with AsCpf1, LbCpf1, and SpCas9 at 10 endogenous sites.** Indel frequencies obtained with AsCpf1, LbCpf1 and SpCas9 at 10 endogenous sites. Cpf1 target sequence and SpCas9 target sequence were designed to be overlapped.

## 2. Analysis of target specificity of Cpf1

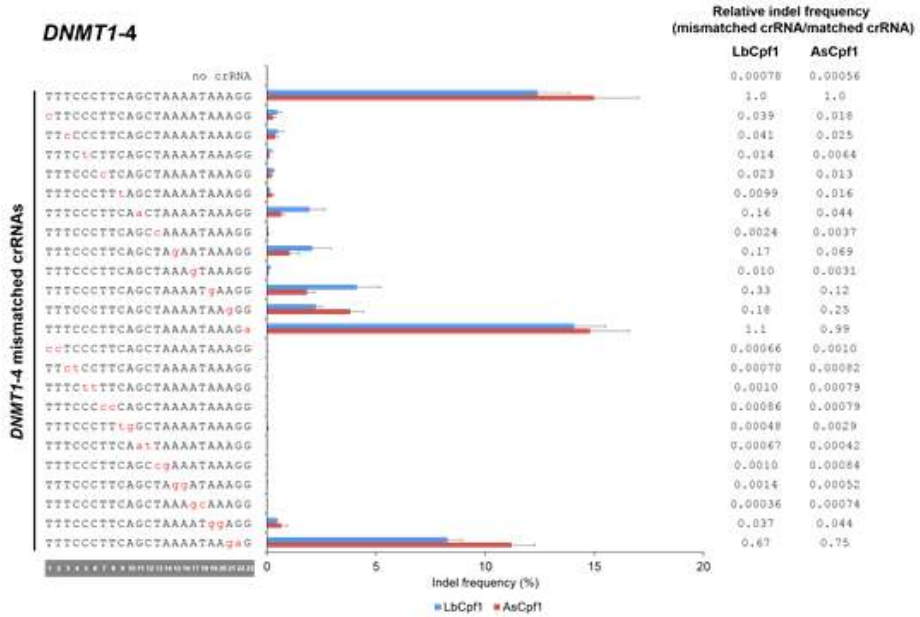
### a. Analysis of target specificity of Cpf1 using mismatched crRNAs

To verify the target specificity of Cpf1 in the context of mammalian genomes, I first investigated the effect of mismatches between guide sequence of the crRNA and target DNA systematically in human cells. AsCpf1 and LbCpf1 recognize and cleave 27-nucleotide target DNA sequences composed of the 5'-TTTN-3' PAM sequence followed by the 23-nucleotide of protospacer sequence complementary to guide sequence of crRNA. I chose three endogenous target sites and carried out cell test with Cpf1 expression plasmids and crRNA-encoding plasmids which contained single or double nucleotide mismatches to the target DNA in the guide sequence of crRNA. The DNA cleavage efficiencies obtained with matched crRNA and mismatched crRNAs were measured by targeted deep sequencing (Figure 20). Strikingly, Cpf1 did distinguish single mismatches at position 1-17 and double mismatches at position 1-18 almost completely abolished DNA cleavage activities of Cpf1 for the DNMT1-4 site (Figure 20b). In general, Cpf1 did tolerate single mismatches, whereas double mismatches at position 1-18 led to a substantial loss of indel rates at three endogenous target sites. These results indicate that Cpf1 system works in human cells in a highly specific manner, although it can tolerate mismatches at the PAM-distal region.

**a**


DNMT1-3
Relative indel frequency
(mismatched crRNA/matched crRNA)
LbCpf1
AsCpf1
DNMT1-3 mismatched crRNAs
no crRNA
CTGATGGTCCATGTCTGTTACTC
tTGATGGTCCATGTCTGTTACTC
CTaATGGTCCATGTCTGTTACTC
CTGAcGGTCCATGTCTGTTACTC
CTGATGaTCCATGTCTGTTACTC
CTGATGGTcCATGTCTGTTACTC
CTGATGGTCCgTGTCTGTTACTC
CTGATGGTCCATaTCTGTTACTC
CTGATGGTCCATGTcTGTTACTC
CTGATGGTCCATGTCTaTTACTC
CTGATGGTCCATGTCTGTcACTC
CTGATGGTCCATGTCTGTTAtTC
CTGATGGTCCATGTCTGTTACTt
tcGATGGTCCATGTCTGTTACTC
CTagTGGTCCATGTCTGTTACTC
CTGAcaGTCCATGTCTGTTACTC
CTGATGacCCATGTCTGTTACTC
CTGATGGTttATGTCTGTTACTC
CTGATGGTCCgcGTCTGTTACTC
CTGATGGTCCATacCTGTTACTC
CTGATGGTCCATGTtcGTTACTC
CTGATGGTCCATGTCTacTACTC
CTGATGGTCCATGTCTGTcgCTC
CTGATGGTCCATGTCTGTTAtcC
0.00015 0.00037
1.0 1.0
0.16 0.33
0.0093 0.063
0.042 0.096
0.050 0.15
0.42 0.51
0.45 0.29
0.18 0.18
0.21 0.19
0.00075 0.011
0.13 0.48
0.99 0.98
1.3 1.2
0.00091 0.021
0.0028 0.0013
0.00019 0.0010
0.0027 0.0027
0.00051 0.027
0.00025 0.0014
0.00018 0.0027
0.00046 0.0012
0.00024 0.00053
0.098 0.39
0.83 0.85
0 10 20 30 40 50 60
Indel frequency (%)
LbCpf1 AsCpf1

**b**


DNMT1-4
Relative indel frequency
(mismatched crRNA/matched crRNA)
LbCpf1
AsCpf1
DNMT1-4 mismatched crRNAs
no crRNA
TTTCCCTTCAGCTAAAATAAAGG
cTTCCCTTCAGCTAAAATAAAGG
TTcCCCTTCAGCTAAAATAAAGG
TTTCtCTTCAGCTAAAATAAAGG
TTTCCCcTCAGCTAAAATAAAGG
TTTCCCTTtAGCTAAAATAAAGG
TTTCCCTTCAaCTAAAATAAAGG
TTTCCCTTCAGCcAAAATAAAGG
TTTCCCTTCAGCTAgAATAAAGG
TTTCCCTTCAGCTAAAgTAAAGG
TTTCCCTTCAGCTAAAATgAAGG
TTTCCCTTCAGCTAAAATAAgGG
TTTCCCTTCAGCTAAAATAAAGa
ccTCCCTTCAGCTAAAATAAAGG
TTatCCTTCAGCTAAAATAAAGG
TTTCttTTCAGCTAAAATAAAGG
TTTCCCccCAGCTAAAATAAAGG
TTTCCCTTtgGCTAAAATAAAGG
TTTCCCTTCAatTAAAATAAAGG
TTTCCCTTCAGCcgAAATAAAGG
TTTCCCTTCAGCTAggATAAAGG
TTTCCCTTCAGCTAAAgcAAAGG
TTTCCCTTCAGCTAAAATggAGG
TTTCCCTTCAGCTAAAATAAgaG
0.00070 0.00056
1.0 1.0
0.039 0.018
0.041 0.025
0.014 0.0064
0.023 0.013
0.0099 0.016
0.16 0.044
0.0024 0.0037
0.17 0.069
0.010 0.0031
0.33 0.12
0.18 0.25
1.1 0.99
0.00066 0.0010
0.00070 0.00082
0.0010 0.00079
0.00086 0.00079
0.00048 0.0029
0.00067 0.00042
0.0010 0.00084
0.0014 0.00052
0.00036 0.00074
0.037 0.044
0.67 0.75
0 5 10 15 20
Indel frequency (%)
LbCpf1 AsCpf1

**c**

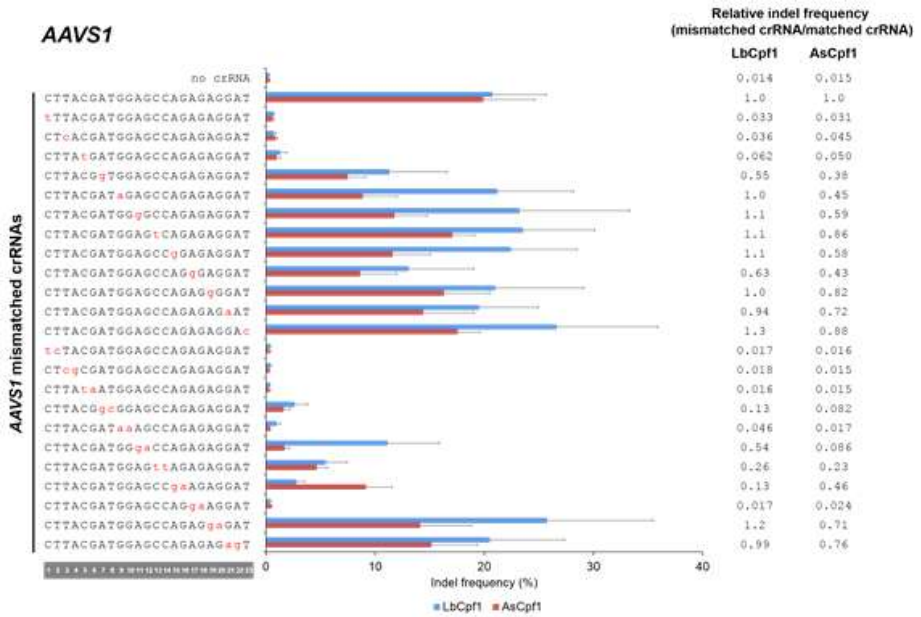



**Figure 20. Analysis of target specificity of Cpf1 with mismatched crRNAs.** Indel frequencies obtained with mismatched crRNAs. Error bars indicate s.e.m. (a) DNMT1-3 target site. (b) DNMT1-4 target site. (c) AAVS1 target site.

## b. Genome-wide target specificity of Cpf1

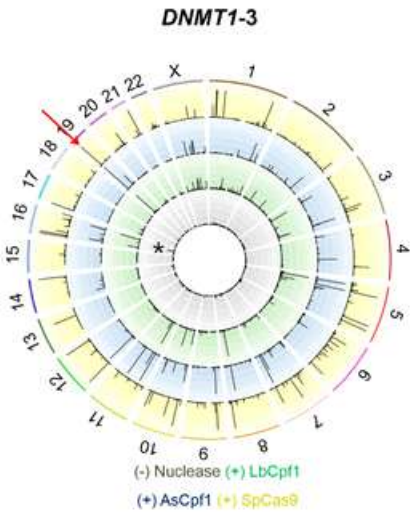
To investigate the target specificity of Cpf1 nucleases in an unbiased manner, Digenome-seq (Kim et al., 2015; Kim et al., 2016b) was performed at the selected sites. Cell-free genomic DNA was digested with preassembled AsCpf1, LbCpf1 and SpCas9 ribonucleoproteins (RNPs) and the digested genomic DNA was subjected to whole-genome sequencing (WGS). In general, WGS data produce random alignment of sequencing reads, but treatment of nucleases into genomic DNA in vitro leads to cutting of the on-target and potential off-target sites in the genome and these cleavages produce uniform alignments of sequencing reads at the cleaved sites. In vitro cleavage sites were computationally identified and DNA cleavage scores were calculated at each position. Cpf1 was highly specific based on the genome-wide Circos plots of DNA cleavage scores, cleaving chromosomal DNA in vitro at just 46 sites for the DNMT1-3 target site and one site for the DNMT1-4 target site (Figure 21). The number of in vitro cleavage sites identified in monoplex Digenome-seq analyses was 12 ± 5 for AsCpf1 or 6 ± 3 for LbCpf1, which is far less than the 90 ± 30 that was calculated with SpCas9 in a previous multiplex Digenome-seq experiment (Kim et al., 2016b) (Figure 22).

Next, Digenome-captured sites were validated with targeted deep sequencing for measuring the DNA cleavage efficiencies at these sites in the context of human genomes and evaluating the sensitivity of Digenome-seq analysis. 46 sites were captured for the DNMT1-3 target

site, and only 4 sites showed the indel rates above basal sequencing errors (Figure 23). Indel rates of the most of the validated off-target sites were below 1%, much lower than those at corresponding on-target sites. Mismatches of the validated off-target sites were in the 3' PAM distal region mostly. These results seem to be correlated with the results of mismatched crRNA test (Figure 20).

Notably, the OT6 off-target site harbored a non-canonical, 5'-TCTN-3' PAM sequence (Figure 23). This result suggests that Cpf1 can cleave chromosomal target sites that contain non-canonical PAMs. To investigate a success rate of Cpf1-mediated genome editing at chromosomal target sites harboring non-canonical PAMs, I selected the target sites which contained non-canonical PAM sequences, such as 5'-TTCN-3' and 5'-TCTN-3' and carried out the cell test. Success rates were calculated based on the number of target sites which showed indel rates above 1%. The success rate of Cpf1 mediated genome editing at the selected sites containing non-canonical PAMs was approximately 50%, while Cpf1 showed 100% of targeting at the sites harboring canonical 5'-TTTN-3' PAM (Figure 24). The targeted sites also showed an overall low indel rates, compared with the indel rates of the target sites harboring canonical PAM sequences. These results indicate that further engineering is necessary to increase the range of Cpf1-mediated genome editing.

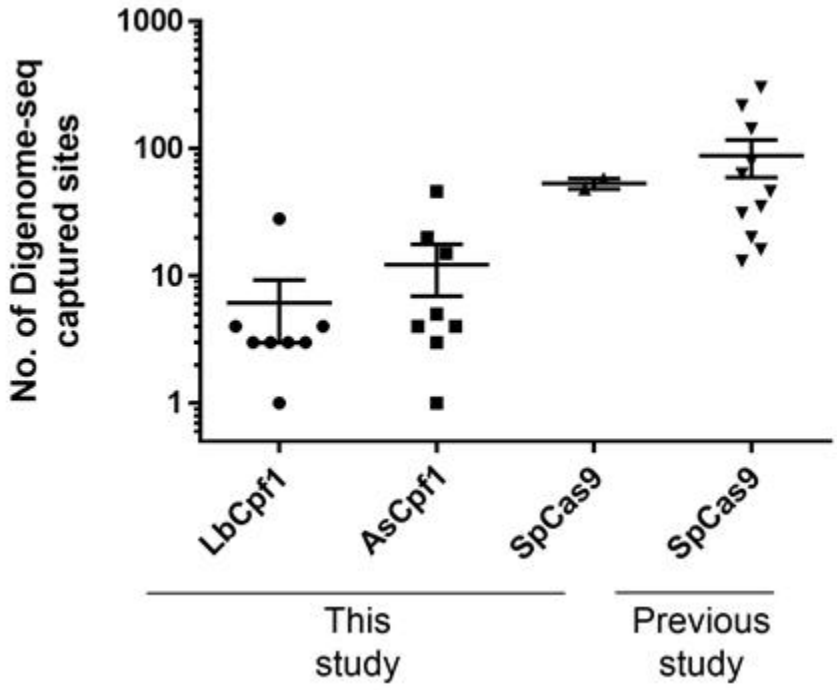
**a**




**b**

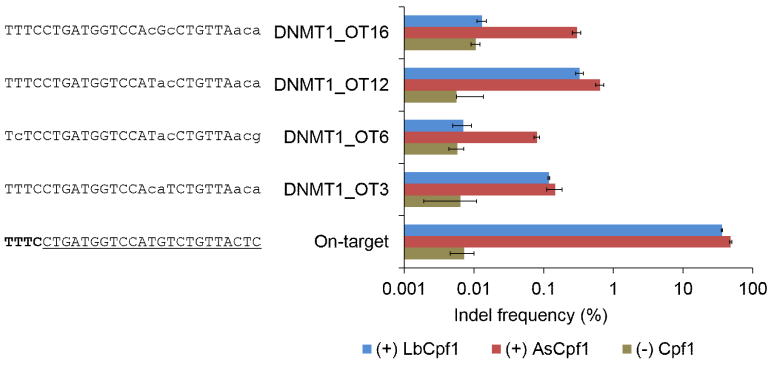

(by Daesik Kim in Seoul National University)

**Figure 21. Genome-wide Circos plots of DNA cleavage scores obtained with AsCpf1, LbCpf1 and SpCas9.** Hela genomic DNA was digested with AsCpf1, LbCf1 and SpCas9 and whole genome sequencing (WGS) was performed for digested genomic DNA samples. In vitro cleavage sites obtained by nuclease treatment were captured through bioinformatic analysis of WGS data and DNA cleavage scores were calculated based on WGS raw data. Genome-wide Circos plots show DNA cleavage sites and DNA cleavage scores for each sample. (a) Genome-wide Circos plots for the DNMT1-3 target site. Red arrow indicates false positive. The asterisk indicates the position of on-target site. (b) Genome-wide Circos plots for the DNMT1-4 target site.




(by Daesik Kim in Seoul National University)

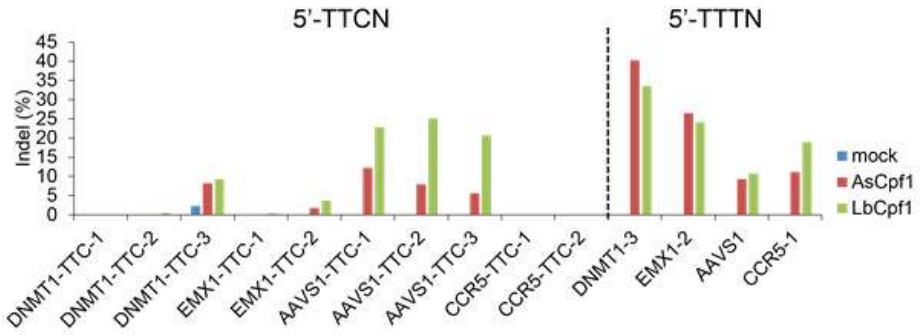
**Figure 22. The number of in vitro cleavage site.** The number of Digenome-seq captured sites were calculated using monoplex Digenome-Seq with LbCpf1 (n=8), AsCpf1 (n=8), and SpCas9 (n=2). Previous multiplex Digenome-Seq data for SpCas9 (n=11) were plotted at the right side.



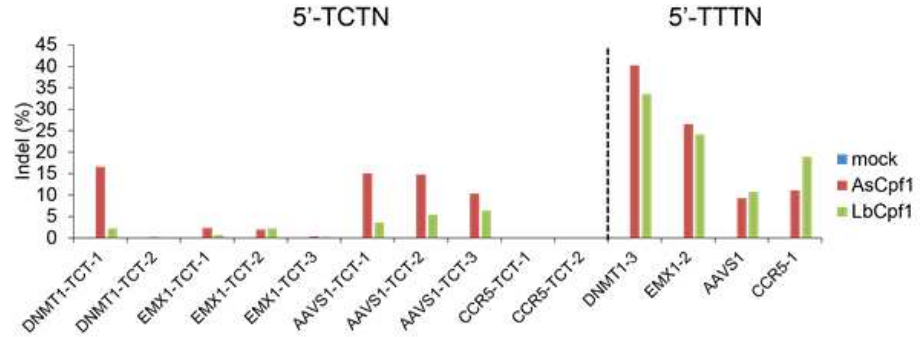

(by Daesik Kim in Seoul National University)

**Figure 23. Validation of in vitro cleavage sites by targeted deep sequencing.** Digenome-captured sites were validated by targeted deep sequencing. Two sites for LbCpf1 and four sites for AsCpf1 were validated with true off-target sites for DNMT1-3 crRNAs. PAM sequence is shown in bold and mismatched nucleotides are shown in lower case. The target sequence is underlined. Error bars indicate s.e.m.

**a**


5'-TTCN
5'-TTTN
Indel (%)
45
40
35
30
25
20
15
10
5
0
mock
AsCpf1
LbCpf1
DNMT1-TTC-1
DNMT1-TTC-2
DNMT1-TTC-3
EMX1-TTC-1
EMX1-TTC-2
AAVS1-TTC-1
AAVS1-TTC-2
AAVS1-TTC-3
CCR5-TTC-1
CCR5-TTC-2
DNMT1-3
EMX1-2
AAVS1
CCR5-1

**b**


5'-TCTN
5'-TTTN
Indel (%)
45
40
35
30
25
20
15
10
5
0
mock
AsCpf1
LbCpf1
DNMT1-TCT-1
DNMT1-TCT-2
EMX1-TCT-1
EMX1-TCT-2
EMX1-TCT-3
AAVS1-TCT-1
AAVS1-TCT-2
AAVS1-TCT-3
CCR5-TCT-1
CCR5-TCT-2
DNMT1-3
EMX1-2
AAVS1
CCR5-1

**c**

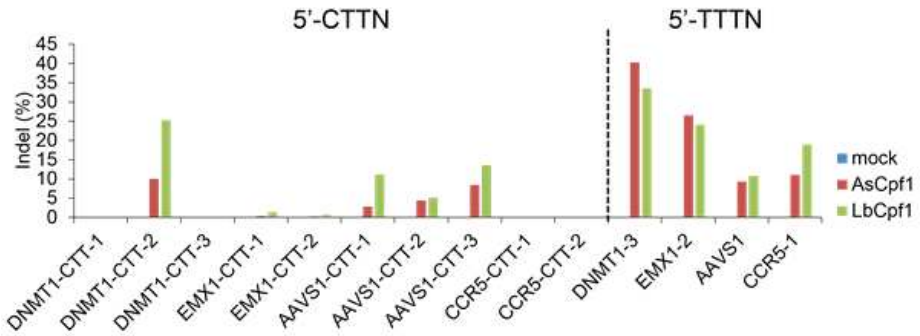
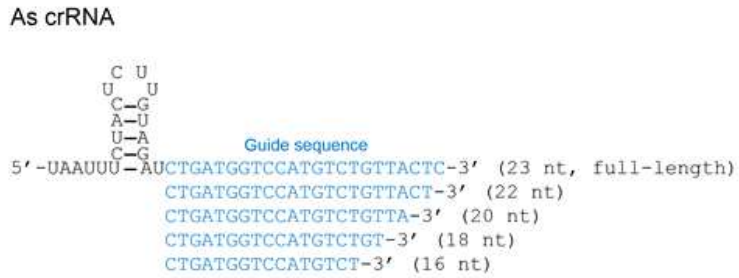

5'-CTTN
5'-TTTN
Indel (%)
45
40
35
30
25
20
15
10
5
0
mock
AsCpf1
LbCpf1
DNMT1-CTT-1
DNMT1-CTT-2
DNMT1-CTT-3
EMX1-CTT-1
EMX1-CTT-2
AAVS1-CTT-1
AAVS1-CTT-2
AAVS1-CTT-3
CCR5-CTT-1
CCR5-CTT-2
DNMT1-3
EMX1-2
AAVS1
CCR5-1

**Figure 24. Cpf1-mediated gene editing at target sites containing non-canonical PAM sequences.** (a) Cpf1-mediated gene editing at 10 endogenous sites containing 5'-TTCN PAM. Indel rates were measured by targeted deep sequencing. (b) Cpf1-mediated gene editing at 10 endogenous sites containing 5'-TCTN PAM. (c) Cpf1-mediated gene editing at 10 endogenous sites containing 5'-CTTN PAM.

## c. Analysis of target specificity of Cpf1 using truncated crRNAs

In the previous study, SpCas9 showed an improvement of specificity of nucleases with truncated gRNAs (Fu et al., 2014). To test whether truncated crRNA also could reduce indel rates of validated off-target sites, I designed a series of truncated crRNAs (Figure 25a) and carried out cell test. Indel rates of the on-target site and the off-target sites were measured by targeted deep sequencing. I found that crRNAs truncated at the 3'end enabled an up to ninefold reduction of indel frequencies at certain off-target sites (Figure 26b). Unfortunately, this approach is limited, as most off-target sites harbor mismatches at the PAM –distal 3' end.

**a**




**b**

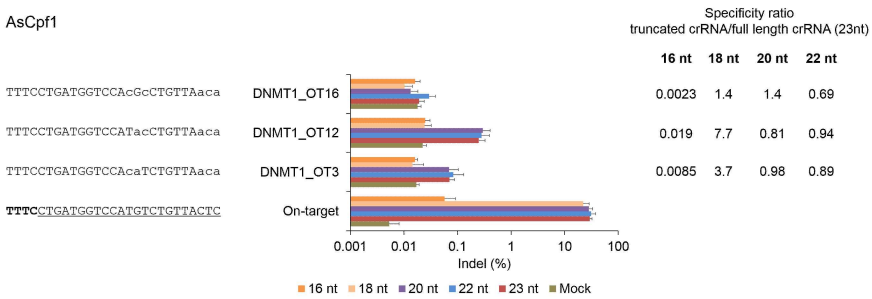



**Figure 25. Analysis of target specificity of Cpf1 with truncated crRNAs.** (a) Design for truncated crRNA. (b) Each truncated crRNA and full-length crRNA for DNMT1-3 target sites were transfected into HEK293T17 cells with AsCpf1-encoding plasmids. Indel rates of on-target site and validated off-target sites were analyzed by targeted deep sequencing. Specificity indicates on-target indel rate over off-target indel rate. Error bars indicate s.e.m. PAM sequence is shown in blue and mismatched nucleotides are shown in red and lower case.

# Discussion

In the recent research, CRISPR-Cpf1 system, a putative type V CRISPR-Cas9 system was identified and this system was successfully repurposed for targeted genome editing in human cells (Zetsche et al., 2015a). Cpf1 system consists of two components – Cpf1 protein and crRNA-, and target specificity of Cpf1 system is determined by 23-nucleotides of guide sequence. These traits make CRISPR-Cpf1 system as a simple platform for targeted genome engineering in human cells, model organisms, and plants like CRISPR-Cas9 system derived from Streptococcus pyogenes. Despite its potentials, target specificities of Cpf1 endonucleases have not been elucidated thoroughly, leaving a question about the off-target issue. To address this question, I optimized the platform for Cpf1-mediated genome editing in human cells and analyzed the target specificity of Cpf1 endonucleases. At the optimized condition, Cpf1 can induce targeted genome modification efficiently and Cpf1 shows a high sensitivity about the mismatches between guide sequence of crRNA and target DNA in human cells, suggesting that Cpf1 endonuclease is suitable for a precise genome editing.

Improving the programmable nucleases for reducing off-target effects is a major goal of the field of genome engineering. In the clonal level, off-target effects of programmable nucleases are usually negligible (Kim et al., 2015; Park et al., 2015; Suzuki et al., 2014), because indel rates of off-target sites are generally lower than those of on-target sites and clone samples that harbor the genetic modifications

at the off-target sites can be easily excluded from a further research by manual handling. But, off-target effects can be a big problem in the case of biomedical applications of these technologies. For examples, ex vivo cell therapy needs a large number of genome-corrected cells for transplantation and a little portion of these cells that might contain the genetic modifications at the off-target sties can yield concerns about the safety issue. Recently, several groups independently reported the rationally engineered Cas9 variants to address this concern (Kleinstiver et al., 2016; Slaymaker et al., 2016). These studies demonstrate that Cas9 variants engineered by modification of the residues of SpCas9 interacting with non-target DNA strand (Slaymaker et al., 2016) or target DNA strand (Kleinstiver et al., 2016) show reduced off-target effects compared to those obtained with wild-type SpCas9 and robust on-target cleavages. Although I did not directly compare the target specificity of Cpf1 endonuclease with that of Cas9 variants in this study, Cpf1 endonuclease has a potential for showing better performance than eSpCas9 and SpCas9-HF. First, mismatched crRNA test showed that double mismatches at position 1-18 led to substantial loss of DNA cleavage activity of wild-type Cpf1 and even single mismatches caused a loss of DNA cleavage activity of Cpf1 for the DNMT1-4 site. Second, Cpf1 generally showed the number of Digenome-captured sites, fewer than being present in wild-type Cas9 nuclease. These results indicate that wild type of Cpf1 endonuclease generally shows a high specificity compared with wild type of SpCas9 and it can be further improved by rational engineering of Cpf1 protein

based on the crystal structure information or delivery of Cpf1 nucleases as a preassembled, recombinant Cpf1 ribonucleoproteins (Dong et al., 2016; Gao et al., 2016; Hur et al., 2016; Kim et al., 2014b; Yamano et al., 2016)

Cpf1-endonucleases recognize the 5'-TTTN-3' as a PAM sequence, which might limit a broad use of this nuclease for correcting mutations related with human diseases because long PAM sequence restricts a design of target sites at the desired site. To harness the potential of Cpf1-endonuclease fully, engineering of PAM specificity of Cpf1 nuclease will be of vital importance in the near future. In recent studies, PAM specificity of SpCas9 and SaCas9 was successfully engineered using a directed evolution and E.coli-based selection system (Kleinstiver et al., 2015a; Kleinstiver et al., 2015b). With this method or other approaches, Cpf1 variants, which recognize the non-canonical PAM efficiently will be identified and expand the utility of this system in gene therapy. I found that wild-type of Cpf1 could recognize and cleave the target sites harboring non-canonical PAMs such as 5'-TTCN-3' and 5'-TCTN-3' although success rate and gene modification efficiency were remarkably low. Improving the success rate and gene modification efficiency of Cpf1 at the target sites that contain these PAM sequences is worth a try in the first place.

Efficient delivery of programmable nucleases is also a major issue in in vivo gene therapy. In case of in vivo delivery, recombinant adeno-associated virus (rAAV) is clinically used as a vector because of low cytotoxicity and low possibility of integrating of viral genome into

the human genome randomly. AAV has a packaging capacity of ~ 4.7 kb, hampering the efficient packaging of Cpf1 expression cassette owing to the relatively large genetic size of Cpf1. Splitting of Cpf1 nuclease might be a good approach for facilitating the packaging of expression cassette and the increasing the efficiency of in vivo genome editing and epi-genome editing (Chew et al., 2016; Truong et al., 2015; Wright et al., 2015; Zetsche et al., 2015b).

In summary, CRISPR-Cpf1 demonstrates efficient genome editing in human cells and shows that this nuclease induces targeted genome modification in a highly specific manner. Thus, CRISPR-Cpf1 will be a good option for RNA-guided genome editing and broaden the usages of this programmable nuclease in basic research, biotechnology, and biomedical research.

# References

Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics 30, 1473-1475.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-1712.

Bibikova, M., Golic, M., Golic, K.G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161, 1169-1175.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509-1512.

Capecchi, M.R. (1989). Altering the genome by homologous recombination. Science 244, 1288-1292.

Chew, W.L., Tabebordbar, M., Cheng, J.K., Mali, P., Wu, E.Y., Ng, A.H., Zhu, K., Wagers, A.J., and Church, G.M. (2016). A multifunctional AAV-CRISPR-Cas9 and its host response. Nature methods 13, 868-874.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013a). Targeted genome engineering in human cells with the Cas9 RNA-guided

endonuclease. Nature biotechnology 31, 230-232.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome research 24, 132-141.

Cho, S.W., Lee, J., Carroll, D., Kim, J.S., and Lee, J. (2013b). Heritable gene knockout in Caenorhabditis elegans by direct injection of Cas9-sgRNA ribonucleoproteins. Genetics 195, 1177-1180.

Choe, S., Schmitz, R.J., Fujioka, S., Takatsuto, S., Lee, M.O., Yoshida, S., Feldmann, K.A., and Tax, F.E. (2002). Arabidopsis brassinosteroid-insensitive dwarf12 mutants are semidominant and defective in a glycogen synthase kinase 3beta-like kinase. Plant physiology 130, 1506-1515.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757-761.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823.

Cradick, T.J., Fine, E.J., Antico, C.J., and Bao, G. (2013). CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. Nucleic acids research 41,

9584-9592.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607.

Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., et al. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. Nature 532, 522-526.

Fedorov, Y., Anderson, E.M., Birmingham, A., Reynolds, A., Karpilow, J., Robinson, K., Leake, D., Marshall, W.S., and Khvorova, A. (2006). Off-target effects by siRNA can induce toxic phenotype. Rna 12, 1188-1196.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nature biotechnology 33, 179-186.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology 31, 822-826.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature biotechnology 32, 279-284.

Gao, P., Yang, H., Rajashankar, K.R., Huang, Z., and Patel, D.J. (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a

unique mechanism for crRNA-mediated target DNA recognition. Cell research 26, 901-913.

Guilinger, J.P., Thompson, D.B., and Liu, D.R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nature biotechnology 32, 577-582.

Hale, C.R., Cocozaki, A., Li, H., Terns, R.M., and Terns, M.P. (2014). Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex. Genes & development 28, 2432-2443.

Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., Steinfeld, I., Lunstad, B.D., Kaiser, R.J., Wilkens, A.B., et al. (2015). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. Nature biotechnology 33, 985-989.

Hess, G.T., Fresard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nature methods.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proceedings of the National Academy of Sciences of the United States of America 111, 6618-6623.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune

system of bacteria and archaea. Science 327, 167-170.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology 31, 827-832.

Hur, J.K., Kim, K., Been, K.W., Baek, G., Ye, S., Hur, J.W., Ryu, S.M., Lee, Y.S., and Kim, J.S. (2016). Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. Nature biotechnology 34, 807-808.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229.

Jackson, A.L., Burchard, J., Schelter, J., Chau, B.N., Cleary, M., Lim, L., and Linsley, P.S. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. Rna 12, 1179-1187.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013).

RNA-programmed genome editing in human cells. eLife 2, e00471.

Jones, H.D. (2015). Regulatory uncertainty over genome editing. Nature plants 1, 14011.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.I., and Kim, J.S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nature methods 12, 237-243, 231 p following 243.

Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H., and Kim, J.S. (2016a). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nature biotechnology 34, 863-868.

Kim, D., Kim, S., Kim, S., Park, J., and Kim, J.S. (2016b). Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. Genome research 26, 406-415.

Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. Nature reviews Genetics 15, 321-334.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W., and Kim, J.S. (2009). Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome research 19, 1279-1288.

Kim, J., and Kim, J.S. (2016). Bypassing GMO regulations with CRISPR gene editing. Nature biotechnology 34, 1014-1015.

Kim, J.M., Kim, D., Kim, S., and Kim, J.S. (2014a). Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. Nature

communications 5, 3157.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014b). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome research 24, 1012-1019.

Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E., et al. (2013). A library of TAL effector nucleases spanning the human genome. Nature biotechnology 31, 251-258.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences of the United States of America 93, 1156-1160.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529, 490-495.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V.V., Zheng, Z., and Joung, J.K. (2015a). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nature biotechnology 33, 1293-1298.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R., et al. (2015b). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523, 481-485.

Koh, S., Lee, S.C., Kim, M.K., Koh, J.H., Lee, S., An, G., Choe, S., and Kim, S.R. (2007). T-DNA tagged knockout mutation of rice OsGSK1, an orthologue of Arabidopsis BIN2, with enhanced tolerance to various abiotic stresses. Plant molecular biology 65, 453-466.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420-424.

Lee, H.J., Kim, E., and Kim, J.S. (2010). Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome research 20, 81-89.

Li, J.F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M., and Sheen, J. (2013). Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nature biotechnology 31, 688-691.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. Nature biotechnology 30, 390-392.

Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nature methods.

Makarova, K.S., and Koonin, E.V. (2015). Annotation and Classification of CRISPR-Cas Systems. Methods in molecular biology 1311,

47-75.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. Nature reviews Microbiology 13, 722-736.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science 339, 823-826.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., et al. (2011). A TALE nuclease architecture for efficient genome editing. Nature biotechnology 29, 143-148.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science 326, 1501.

Mulepati, S., Heroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science 345, 1479-1484.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D., and Kamoun, S. (2013). Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nature biotechnology 31, 691-693.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and

vertebrate adaptive immune systems. Science 353.

Park, C.Y., Kim, D.H., Son, J.S., Sung, J.J., Lee, J., Bae, S., Kim, J.H., Kim, D.W., and Kim, J.S. (2015). Functional Correction of Large Factor VIII Gene Chromosomal Inversions in Hemophilia A Patient-Derived iPSCs Using CRISPR-Cas9. Cell stem cell 17, 213-220.

Pennisi, E. (2013). The CRISPR craze. Science 341, 833-836.

Pennisi, E. (2016). The plant engineer. Science 353, 1220-1224.

Rahdar, M., McMahon, M.A., Prakash, T.P., Swayze, E.E., Bennett, C.F., and Cleveland, D.W. (2015). Synthetic CRISPR RNA-Cas9-guided genome editing in human cells. Proceedings of the National Academy of Sciences of the United States of America 112, E7110-7117.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186-191.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.L., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. Nature biotechnology 31, 686-688.

Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell research 18, 134-147.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and

Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science 351, 84-88.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nature reviews Microbiology 6, 181-186.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annual review of biochemistry 82, 237-266.

Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., et al. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. Molecular cell 56, 518-530.

Sung, Y.H., Kim, J.M., Kim, H.T., Lee, J., Jeon, J., Jin, Y., Choi, J.H., Ban, Y.H., Ha, S.J., Kim, C.H., et al. (2014). Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome research 24, 125-131.

Suzuki, K., Yu, C., Qu, J., Li, M., Yao, X., Yuan, T., Goebl, A., Tang, S., Ren, R., Aizawa, E., et al. (2014). Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. Cell stem cell 15, 31-36.

Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. Current opinion in microbiology 14, 321-327.

Truong, D.J., Kuhner, K., Kuhn, R., Werfel, S., Engelhardt, S., Wurst, W., and Ortiz, O. (2015). Development of an intein-mediated split-Cas9 system for gene therapy. Nucleic acids research 43, 6450-6458.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nature biotechnology 32, 569-576.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature biotechnology 33, 187-197.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature 435, 646-651.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nature reviews Genetics 11, 636-646.

Waltz, E. (2016). Gene-edited CRISPR mushroom escapes US regulation. Nature 532, 293.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice

carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153, 910-918.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature 477, 486-489.

Woo, J.W., Kim, J., Kwon, S.I., Corvalan, C., Cho, S.W., Kim, H., Kim, S.G., Kim, S.T., Choe, S., and Kim, J.S. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nature biotechnology 33, 1162-1164.

Wright, A.V., Sternberg, S.H., Taylor, D.W., Staahl, B.T., Bardales, J.A., Kornfeld, J.E., and Doudna, J.A. (2015). Rational design of a split-Cas9 enzyme complex. Proceedings of the National Academy of Sciences of the United States of America 112, 2984-2989.

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., et al. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell 165, 949-962.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015a). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759-771.

Zetsche, B., Volz, S.E., and Zhang, F. (2015b). A split-Cas9 architecture for inducible genome editing and transcription modulation. Nature biotechnology 33, 139-142.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X., Pierick, C.J., Dobbs, D., Peterson, T., et al. (2010). High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proceedings of the National Academy of Sciences of the United States of America 107, 12028-12033.

Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker, C.G., Bogdanove, A.J., and Voytas, D.F. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant physiology 161, 20-27.

# 국문 초록

유전자 가위를 이용한 유전체 교정은 생명 과학 연구 분야에서 필수적인 기술이다. 가장 많이 사용되고 있는 유전자 가위로 zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), RNA-guided engineered nucleases (RGENs)이 있고, 그중에서 RGENs은 guide RNA의 서열을 바꾸는 것으로 표적을 자유롭게 결정할 수 있는 편리성으로 최근 주목받고 있다.

이번 연구에서는 Cas9 단백질과 guide RNA 복합체를 식물 원형질체에 polyethylene glycol (PEG)를 이용해 직접 전달하여 높은 효율로 식물 유전체를 교정하였고, 교정된 세포로부터 개체를 확보할 수 있었다. 해당 방식은 plasmid DNA 대신 세포 내에서 빨리 작동하고 분해되는 Cas9 단백질과 guide RNA 복합체를 사용했기 때문에 RGENs의 오작동 가능성을 줄일 수 있었고, 유전자 교정 식물 관련 규제로부터 벗어날 수 있는 가능성이 높을 것으로 기대된다. 또한, 최근에 새롭게 보고된 CRISPR-Cpf1 시스템의 표적 특이성을 분석하였다. CRISPR-Cpf1 시스템은 인간 배양세포에서 효율적으로 유전체 교정을 수행하면서, 매우 특이적으로 표적 위치에 DNA DSBs를 일으키는 것을 확인할 수 있었다. 이러한 결과를 바탕으로 CRISPR-Cpf1이 매우 정교한 유전자 가위 형태임을 확인할 수 있었다. CRISPR-Cpf1은 기존의 유전자 가위 기술이 표적할 수 있는 영역을 확장시키는 데 중요한 역할을 할 것으로 기대된다.

**학 번:** 2012-23038